Interrogating telomerase activity and telomere function through in-flask template evolution

By

© 2017

Margaret R. Pruitt

B.S., Duke University, 2009

Submitted to the graduate degree program in Molecular and Integrative Physiology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

co-Chair: Peter Baumann, Ph.D.

_____

Fariba Behbod, Pharm.D., Ph.D.

_____

Joan Conaway, Ph.D.

_____

Timothy Fields, M.D., Ph.D.

_____

Scott Hawley, Ph.D.

_____

Brenda Rongish, Ph.D.

_____

Pamela V. Tran, Ph.D.

_____

co-Chair: Michael W. Wolfe, Ph.D.

Date Defended: 11 May 2017

The dissertation committee for Margaret R. Pruitt certifies that this is the approved version of the following dissertation:

Interrogating telomerase activity and telomere function through in-flask template evolution

co-Chair: Peter Baumann, Ph.D.

co-Chair: Michael W. Wolfe, Ph.D.

Date Approved: 11 May 2017

Abstract

The study of telomere evolution has revealed defining features of telomere biology and ultimately has led to the implication of telomeres in cancer and many other human diseases. Telomeres are the nucleoprotein structures at the chromosome end that influence replicative potential, cell viability, and genomic stability. Telomerase is a specialized reverse transcriptase that copies DNA repeats onto the chromosomal 3' overhang from a template region within a stably associated RNA subunit. Dynamic interactions among the telomerase catalytic subunit, the telomerase RNA subunit, telomeric DNA, and sequence-specific, telomere-capping proteins govern telomere maintenance. The research described here investigates which interactions may exert the greatest influence on telomere sequence evolution and how these interactions in turn affect telomerase activity. Additionally, most cancers evade replicative senescence through telomerase-dependent, telomere elongation, making both telomerase and telomeres potential targets for anti-cancer therapy. Telomere destabilization represents one therapeutic strategy where increased telomerase activity in cancer cells can be harnessed to incorporate mutant repeats that impair telomere homeostasis and cause apoptosis. However, the innovation of such therapies requires a better understanding of which telomerase RNA templates can drive the incorporation of mutant repeats at the chromosome end and instigate an immediate apoptotic response. Human telomerase and telomere-associated proteins are structurally and functionally conserved in the fast-growing, genetically-tractable fission yeast, *Schizosaccharomyces pombe*, making it an ideal model organism. This dissertation research employs a prospective evolution approach in *S. pombe* to investigate the role of the telomerase RNA template in telomerase activity and telomere function. The study of seven-nucleotide templates that maintained a competitive growth phenotype demonstrated a consistent five-nucleotide core sequence with a

flexible two-nucleotide sequence. Variation in the sequence and location of these two nucleotides affected telomerase alignment, nucleotide addition, and repeat addition. Two variant template strains were found to shift the alignment region three to four nucleotides from the wild type region to facilitate templated, nucleotide addition, revealing remarkable plasticity in the interaction among the telomerase reverse transcriptase and RNA subunits and the telomere. The study of templates unable to maintain growth in liquid culture revealed a six-nucleotide pattern that may result in telomere destabilization and cellular senescence in fission yeast. These findings can inform strategies to destabilize telomeres and ultimately elicit apoptosis in cancer cells.

This dissertation is dedicated to my parents:

Rose E. Pruitt and the late, Willie H. Pruitt, Jr.

Words can only begin to express my deep gratitude for the love you have shown me. The achievement of this milestone was made possible by the early foundation you provided.

Mom, thank you for building upon that foundation. You have dedicated your life's efforts to my academic success, health, and happiness. You have listened to me and guided me as I developed my personal interests and career aspirations. Thank you for each value you have instilled in me, including a love of learning. Mom, thank you especially for inspiring me to pursue a career that is built on service to others.

Acknowledgements

Countless family members, teachers, administrators, and friends have supported me throughout this academic journey. Although it is impossible to adequately thank each person, I would like to extend a heartfelt thank you to everyone who has encouraged me in his or her own unique way.

I would like to especially recognize Peter Baumann, my graduate research mentor. Peter, as your first MD-PhD student, I appreciate that you welcomed me into your lab and valued my perspective. By training me as a basic scientist, you have provided me with skills that will prove invaluable as I pursue future research endeavors. Your mentorship has helped me sharpen my scientific and grant writing skills. You have spent time discussing my research with me, helping me become a more critical thinker. And you have encouraged me to approach independence in the lab while investing resources into my ideas. As I return to medical school, I am especially grateful for a strong scientific knowledge base. Peter, thank you for broadening my in-depth understanding of model organisms, telomere biology, and the many ways in which telomeres and telomerase affect human health.

To my committee members, Fariba Behbod, Joan Conaway, Tim Fields, Scott Hawley, Brenda Rongish, Pam Tran, and Michael W. Wolfe: it has been an honor to learn from each of you. I am deeply grateful for the feedback, collaboration, and support you all have provided throughout my predoctoral training.

Morgan Schroeder, you are an intelligent and motivated researcher. It truly has been a pleasure to collaborate with you and assist your growth as a bioinformaticist and biologist.

Tim Fields and Brenda Rongish, as the Director and Associate Director of the KUMC MD-PhD Program, respectively, you have set the standard for leadership of an outstanding academic program. I am indebted to you for granting me the opportunity and resources to pursue my life-long dream of becoming a physician scientist. Janice Fletcher, as the Administrative Manager, you made it easier for me to worry less about logistics and focus more on studying and research. Thank you!

One of my greatest fortunes has been the presence of exceptional mentors early in my research career. I would like to especially acknowledge Moses Williams, Director of the Physician Scientist Training Program (PSTP) summer internship at Temple University. As a summer scholar in PSTP, I was introduced to the possibility of an MD, PhD degree, placed in my first research lab experience, and connected to peers who were also interested in research and medicine, many of whom I keep in touch with to this day. Moses, thank you for such a grand opportunity. I will always strive to be in “teaching mode”!

Art Balliet, you welcomed me as a high school student into the lab and taught me good wet bench technique. Thank you for taking a chance on such a young student.

Bev Mock, as one of the first principle investigators to invest in my future, you have provided me with my longest research experience prior to beginning graduate school. Know

that I am grateful for the innumerable ways in which you have nurtured my scientific curiosity and supported my career aspirations.

Alan G. Hinnebusch, Edward Levin, Shuling Zhang, and Doug Lowy each of your mentorship has been invaluable throughout my training.

Being part of a vibrant and collaborative academic community is essential to the intellectual development, personal growth, and research success of a PhD student. Thank you to the University of Kansas Medical Center and Stowers Institute for Medical Research colleagues and staff who have lent their support. To all of the Baumann Lab members, thank you for making the lab such an enjoyable place to learn and discover. Lili Pan, Rachel Helston, Diego Paez-Moscoso, Aracely Newton, Lisa Lassise, and Katie Hildebrand thank you especially. Kim Tu! Thank you for your constant encouragement. Naomi Tjaden Butler and Danny Miller, I appreciate your advice and friendship. Stacey Hanlon, Joaquín Navajas Acedo, Diego, Lili, and Shriram Venkatesan thank you for helpful discussion as I prepared for the defense presentation. Asona Lui-Chacon and Carrie Malcom, I am delighted and thankful that together we embarked upon this eight year training toward the dual degree. We're almost there!

To my longest-standing study buddy, my brother, William H. Pruitt, IIII, thank you for believing in me. Know that you have not only inspired me, but also taught me, to be a better overall writer and close-reader. Thankfully, we could visit and talk despite the distance. Come through Video Chat!

I must acknowledge all of my ancestors and extended family members. You who have made it possible for me to pursue this degree. Most importantly, to my grandparents, Matilda and Joseph Chase and Mell and Willie H. Pruitt, Sr.: I am profoundly grateful for the love you have shown me and the examples each of you has set.

The Stowers Institute is an apt establishment of Jim and Virginia Stowers' legacy. It has been a privilege to train at such a premier institution.

Table of Contents

List of Tables

Chapter II

Chapter III

List of Figures

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter I: Introduction

I.1 Introduction to telomeres

Cell-based life forms can be composed of a single cell to trillions of cells, which carry out molecular processes. These molecular processes are important determinants of cellular function and dysfunction and ultimately affect the health or disease of the whole organism. Instructions for these infinite molecular processes are encoded in the genome of an organism. The genome is composed of deoxyribonucleic acid (DNA) sequences organized into chromosomes, which are linear in eukaryotes. Because linear chromosomes have free ends, they present two significant challenges for cells: the end protection problem and the end replication problem. A specialized type of DNA called the telomere and the enzyme that synthesizes it, telomerase, provide a solution for most eukaryotes.

I.1.1 The end protection problem

DNA constantly encounters mutagenic agents within the nuclear environment of the cell. More specifically, exposure to such agents including radiation, reactive oxygen species, and chemicals can damage DNA and compromise its stability. Double strand breaks (DSBs) are one type of DNA damage and result from two nicks in the ribose-phosphate backbone of both strands of the DNA double helix. DSBs can be particularly harmful to cells causing stalled DNA replication forks, chromosomal missegregation, chromosomal rearrangements, and even cell death if left unrepaired or repaired incorrectly [1]. DSBs can be repaired through several mechanisms. One mechanism, classical non-homologous end-joining (c-NHEJ), requires the orchestration of proteins that recognize the damaged site, prepare the site for repair, and directly fuse the ends of the break. A second mechanism, homologous recombination (HR) also requires damage recognition and preparation but alternatively uses a template to regenerate a resected

DNA region around the break [2]. The free ends of a linear chromosome then closely resemble the sites of a DSB which necessitate repair. Yet, the recognition and repair of a natural chromosome end as a DSB have grave consequences for the cell including genomic instability caused by chromosome end to end fusions [3].

Although the tools to understand the molecular mechanisms of DNA damage repair were not yet available, Hermann Muller and Barbara McClintock, in the 1930s and 1940s, reported critical observations regarding the structural changes of chromosomes following a DNA damage event. In his study of X-ray treated *Drosophila* cells, Muller noted that specialized structures at chromosome termini were required for the stable propagation of chromosomes in mitosis and meiosis [4]. McClintock's experiments in maize demonstrated the "bridge-breakage-fusion-bridge" cycle, where two broken chromosomes are joined at the sites of breakage forming a bridge that is later severed during telophase to reproduce two broken ends [5]. Importantly, she demonstrated that when broken ends were healed the cycle was discontinued and the ends were permanently protected from fusions, a behavior previously attributed to natural ends [6]. Muller and McClintock formulated the concept of an essential structure at the chromosome end that protected natural ends from DNA damage and thereby distinguished them from true sites of DSBs. Muller and colleagues appropriately termed this structure the "telomere" from the Greek origins "telos" (end) and "meros" (part) [7].

I.1.2 The end replication problem

In the early 1950s, James Watson and Francis Crick proposed a double helix structure for DNA [8]. The advent of such a structure stimulated the postulation of several theories regarding DNA replication. However, only one theory, also proposed by Watson and Crick and later experimentally confirmed by Meselson and Stahl, and independently by Taylor et al. would form

the basis of our present day understanding of semiconservative replication [9-11]. Semiconservative replication is the separation of the double helix into two parent strands that each serve as a template for the synthesis of a daughter strand. Later in the 1950s, Kornberg and colleagues went on to purify a daughter strand synthesizing enzyme, DNA polymerase I, from *E. coli* and determined that, just as with DNA purified from a cellular source, the isolated enzyme generated a 3'-5' phosphodiester bond between the end of the growing strand and the newly added nucleotide [12, 13].

The discovery of semiconservative replication and DNA polymerase led scientists to consider how the chromosome end might be replicated (Figure 1.1). During replication, the double helix is unwound and the two strands become unzipped in a replication fork allowing two strands of replicating DNA. The leading strand grows in the direction of the replication fork whereas the lagging strand is synthesized in the opposite direction [14]. DNA polymerase requires a 3'-hydroxyl (3'-OH) group for the addition of new nucleotides. Therefore, DNA synthesis can only proceed in the 5' to 3' direction and a primer must be provided to initiate DNA synthesis. The primer is an oligonucleotide sequence consisting mostly of ribonucleic acid (RNA) and is later removed from the daughter strand. Because the lagging strand travels in the opposite direction of the replication fork, it is initially synthesized as discontinuous units, called Okazaki fragments [15, 16], that each begin with a primer. As primers are removed, they are replaced by extending the 3' end of the previous Okazaki fragment. However, the removal of the first primer at the 5' most end of the lagging strand cannot be replaced and consequently, a progressive loss of genetic material with each replicative cycle ensues. Alexey Olovnikov and James Watson arrived at the realization of this issue in end replication around the same time [17, 18].

Just a few years earlier, in the early to mid-1960s, Leonard Hayflick and Paul Moorhead had demonstrated that normal human fibroblasts cultured *in vitro* underwent a finite number of divisions (50 ± 10) before becoming senescent [19, 20]. This phenomenon known as, "Hayflick's limit" indicated a biological clock that counted the number of cell divisions and could provide a basis for cellular aging [21]. Additionally, a link between the end replication problem and Hayflick's limit was offered by Olovnikov, who suggested that Muller's "telogenes" (later renamed telomeres) might provide a buffer that progressively shortened thereby limiting the number of cell divisions [22]. He correctly reasoned that the loss of these "telogenes" would cause chromosomal aberrations and cell death.

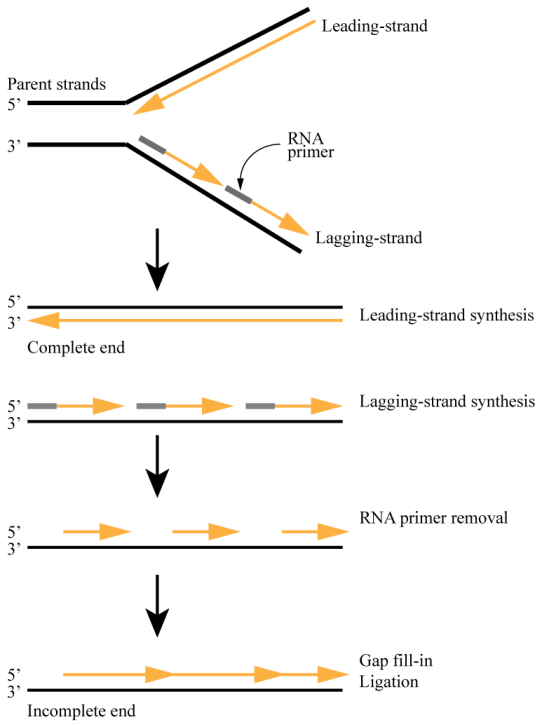


Figure 1.1 The end replication problem. Semiconservative replication involves two daughter strands, the leading-strand and the lagging-strand. A mostly RNA oligonucleotide primer is required to initiate DNA polymerization. Later during replication, RNA primers are removed. Individual fragments of the lagging daughter strand are then extended and ligated to fill in the gaps created by RNA primer removal. However, the 5' most end of the lagging strand remains shortened, resulting in the gradual loss of genetic material with each replicative cycle.

I.2 The evolution of telomerase and telomere sequences

Eukaryotes and biological agents, such as viruses with linear chromosomes, display diverse solutions to the end replication and end protection problems. However, the vast majority of eukaryotes rely on telomeres, the repetitive DNA sequences at the chromosome end bound by regulatory and protective proteins [23, 24]. With rare exceptions, including flies and mosquitos, telomere sequences are elongated by the reverse transcriptase activity of telomerase. In the mid-1970s, equipped with the discoveries of the first three-fourths of the century, researchers began to focus on uncovering these molecular details of telomere biology. If telomeres were specialized DNA structures at the chromosome end, what might be the DNA sequence? Also, since telomeres are essential for cellular division, how might they be replenished, or elongated?

I.2.1 Diverse solutions to the end replication problem

Some bacteriophages, such as phage lambda, solve the end replication problem by circularizing their linear DNA during replication [25]. Phage T7 concatenates its viral genome for replication and later divides the concatemer into smaller, complete linear chromosomes [17]. Alternatively, *Drosophila* maintains linear chromosome ends as telomeres. However, in lieu of telomerase, *Drosophila* makes use of retrotransposable elements that are reverse transcribed onto the chromosome end in long, tandem arrays [26].

I.2.2 The discovery of telomerase and its functional conservation

In the mid-1970s, Elizabeth Blackburn set out to determine the sequence of ribosomal DNA (rDNA) and in so doing discovered the first reported telomeric DNA repeats. She began her study in *Tetrahymena thermophila* owing to its high copy number of short, linear rDNA minichromosomes that could be purified [27]. Using *in vitro* labelling, Blackburn defined the sequence of the minichromosome termini as repeats of CCCCAA motifs organized into

approximately 20 to 70 tandem repeats [28]. She noted that the complementary strand was G-rich, made up of TTGGGG motifs, and ended on either side of the chromosome as an overhang with a 3'-OH. In collaboration with Jack Szostak, Blackburn went on to demonstrate that when *Tetrahymena* telomeric DNA was cloned into a linearized yeast plasmid and transformed into budding yeast, the plasmid was stabilized in its linear form and acquired telomeric yeast repeats [29, 30]. Conversely, in the absence of the *Tetrahymena* telomeric DNA, the linearized yeast plasmid reverted to a circular form. These findings not only revealed the TG_{1-3} repeat sequence of budding yeast telomeres but also signified a conservation in telomere maintenance mechanisms [27]. As a graduate student in the Blackburn lab, Carol Greider provided evidence of terminal transferase activity in *Tetrahymena* cell extracts incubated with radiolabeled deoxy-guanine triphosphates and an oligonucleotide primer resembling the 3' ends of either the *Tetrahymena* or yeast telomeres [31]. She later showed that the telomerase enzyme contained an RNA component that was required for enzymatic activity and carried the reverse complement of the G-rich repeat [32, 33].

Since these initial findings, a wealth of research has revealed the remarkable functional conservation of telomerase despite the compositional diversity of telomerase ribonucleoprotein complexes. The telomerase catalytic subunit contains a conserved reverse transcriptase domain in ciliated protozoa, such as *E. aediculatus* and *T. thermophila*, yeast, including *S. cerevisiae* and *S. pombe*, and vertebrates, including human [24, 34-36]. However, accessory proteins are also required for activity *in vivo* and can vary greatly among vertebrate, fungal, and ciliate telomerases [37]. For example, while the ever-shorter telomere 1 (Est1) protein is part of the holoenzyme in yeast, it is not known to be part of the vertebrate holoenzyme. Other accessory factors share similar functions such as the telomerase RNA stabilizing factors, Sm, Lsm, and

H/ACA box proteins in budding yeast, fission yeast, and vertebrates, respectively. Similarly, the internal telomerase RNA subunit is required for telomerase activity, structural stability, and stipulation of the repeat sequence among eukaryotes that use telomerase [34]. Its sequence and secondary structure are similar among vertebrates [38]. Beyond vertebrates, the telomerase RNA primary sequence can be highly divergent, ranging from ~145 nucleotides long in some ciliates to >2400 nucleotides long in fungi [39-43]. Yet, the function of specific telomerase RNA secondary structures are also conserved in ciliates, budding yeast, and vertebrates [37]. Conservation of some of these structures has also been confirmed in fission yeast. Examples include: 1) the template, which is reverse transcribed as the DNA repeat, 2) the boundary element, which determines the length of a single repeat, and 3) the stem terminus element, which binds the catalytic subunit [40, 44-47]. Whether the pseudoknot, a folding structure important for catalytic subunit binding in ciliates, budding yeast, and vertebrates, also exists in *S. pombe* has yet to be confirmed. However, conservation of a pseudoknot-like nucleotide sequence and the requirement of this sequence for *in vitro* telomerase activity has been demonstrated [43].

I.2.3 Telomeric DNA sequences

Telomerase uses a template region within the RNA subunit to align to the chromosome 3' overhang and reverse transcribe DNA sequences. Just as the length and primary sequence of entire telomerase RNAs are varied, the short template region can also vary leading to diverse repeat sequences. Nevertheless, key features exist among eukaryotes. Telomeres of vertebrates, fungi, plants, protozoa, and most invertebrates are guanine-rich (G-rich), repetitive sequences, whose lengths are maintained within a species-specific range [23]. Additionally, telomeres are double-stranded proximally with a single stranded 3'-overhang on the G-rich strand.

Vertebrate telomeres consist of TTAGGG repeats [48, 49]. A permutation of this sequence, GGTTAG, more accurately represents how a single repeat motif is reverse transcribed from the telomerase RNA template and how it is bound by the single-strand DNA telomere binding protein, Pot1 [50]. The entire 11-nucleotide template contains a pentanucleotide alignment region and a hexanucleotide sequence that is reverse transcribed [39, 41]. Human telomeres are 5-15kb including a 20-400nt 3' overhang [51-55]. Some populations of inbred laboratory mice have been found to harbor much longer telomeres, ranging from 30 to 150kb [56]. Like *Tetrahymena* telomerase, vertebrate telomerase is processive, generating long stretches of consecutive repeats and nearly perfectly homogeneous telomere sequences [57, 58].

Contrastingly, telomeric sequences are especially diverse among fungal species. In the genus *Candida* alone, repeats can range from 23 nucleotides in *Candida albicans* to 8 nucleotides in *Candida guillermondii* [59]. Other budding yeast, such as those found in the genus *Saccharomyces*, generate highly heterogeneous telomeres. For example, the common baker's yeast, *Saccharomyces cerevisiae*, produces repeats with a variable number of guanines, denoted as $TG_{1\text{-}3}$ [30]. Similarly, the fission yeast, *Schizosaccharomyces pombe*, incorporates a core GGTTAC sequence with a variable number of guanines, adenines, and cytosines into each repeat as indicated by $GG_{1\text{-}6}TTACA_{0\text{-}1}C_{0\text{-}1}$ [46, 60, 61]. *S. pombe* telomeres are flanked upstream by a ~19kb subtelomeric sequence (formerly, "telomere-associated sequences or TAS") followed by 24 nucleotides of noncanonical repeats [60, 62]. The *S. pombe* telomere ranges between 70 to ~350nt long [61, 63].

I.3 Telomere-binding proteins

Telomeric repeats serve as a scaffold for the binding of protective and regulatory proteins (Figure 1.2). In mammals, a six-protein complex, called shelterin, inhibits the progression of

DNA damage responses at the chromosome end and mediates telomerase recruitment [63]. In this way, the shelterin complex helps to resolve both the end protection problem and the end replication problem. The conservation of shelterin proteins in yeast has made yeast an invaluable tool in understanding the molecular mechanisms of telomere capping and length maintenance.

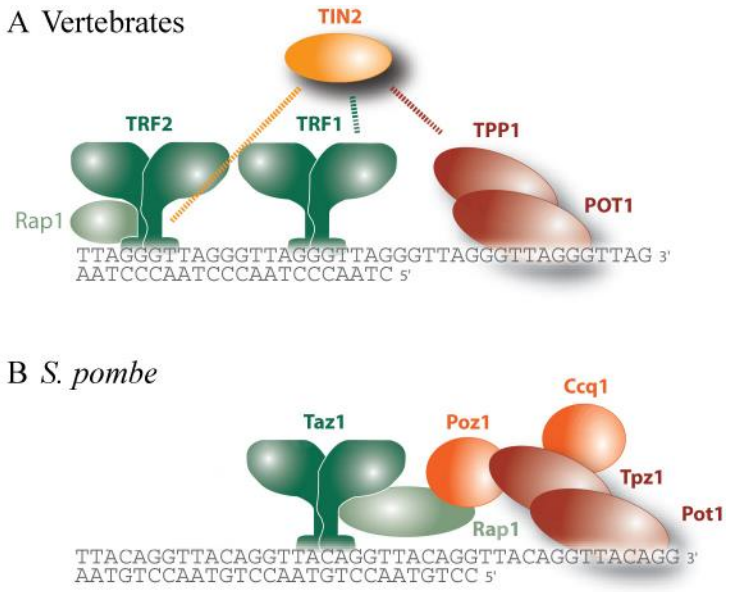


Figure 1.2 Conservation of telomere capping proteins. The regulatory and protective proteins that bind the telomeric repeats are shown. Orthologs are represented by the same color. Together with Rap1, TRF2/Taz1 forms the double-stranded telomeric DNA portion of the capping complex. Pot1 and TPP1/Tpz1 form the 3' overhang, single-stranded telomeric DNA portion of the capping complex. TIN2/Poz1 provide a bridge between the double and single stranded telomeric DNA sub-complexes. In *S. pombe*, Ccq1 additionally associates with the telomere to facilitate telomerase recruitment to the chromosome end. Adapted from [63].

I.3.1 Taz1, the double-stranded DNA binding, telomere associated protein

The telomere-associated in *Schizosaccharomyces pombe* protein (Taz1) was identified using a one hybrid screen of *S. pombe* cDNAs in a strain carrying *S. pombe* telomeric sequences upstream of the GAL4-activatable promoter of a *lacZ* gene [64]. As the ortholog of human telomere repeat factors 1 and 2 (hTRF1 and hTRF2), Taz1 binds the double-stranded part of the

telomere via a helix-turn-helix motif homologous to the hTRF Myb DNA binding domain [64-67]. Taz1 associates with sequences containing three consecutive guanines or three sets of two consecutive guanines separated by TTACA [64]. The conserved Rap1 binding motifs (RBM) of Taz1 and hTRF2 mediate their interactions with Rap1 (Repressor activator protein 1) to form the part of the telomere capping complex, or shelterin [68-70].

I.3.2 Pot1, the single-stranded DNA binding, telomere associated protein

A comparison of *S. pombe* genomic DNA to the genes of known ciliate telomere end binding protein- alpha (TEBP-α) subunits led to the discovery of the protection of telomeres gene, *Pot1* [71]. In the same study, the sequence of the newly identified Pot1 protein was used in the Basic Local Alignment Search Tool (BLAST) to identify the human ortholog, hPot1. In both *S. pombe* and human, two oligonucleotide/oligosaccharide binding (OB) folds are responsible for the specific recognition of single stranded-telomeric DNA by Pot1 [50, 61, 72-75]. In human, Pot1 binds the protein formerly known as TINT1, PTOP, or PIP1 (TPP1) to form the single-stranded part of shelterin [76]. The single stranded-DNA and the double-stranded DNA subcomplexes are connected by the TRF2- and TRF1- Interacting Nuclear protein 2 (TIN2) [77-79]. Similarly, in *S. pombe*, Pot1 binds the TPP1 ortholog in *Schizosaccharomyces pombe* protein (Tpz1) and is linked to the double-stranded DNA subcomplex, Taz1-Rap1, through a fifth protein, Pot1-associated in *Schizosaccharomyces pombe* (Poz1) [80].

Extensive research has investigated how *S. pombe* Pot1 (SpPot1) recognizes repeats of highly heterogeneous *S. pombe* telomeres. During the initial discovery of SpPot1, *Baumann, P. and T.R. Cech.* showed that the protein could bind single-stranded, G-strand DNA but not single-stranded C-strand DNA or duplex DNA [71]. The amino terminal fragment of Pot1 (Pot1Np) was found to bind a minimum of one G**GTT**AC sequence, where the bolded and underlined

nucleotides were most important for sequence specificity [72]. *Lei et al.* later solved the crystal structure of Pot1Np complexed with single-stranded GGTTAC or GGTTA DNA, providing evidence that the sequence specificity was afforded by DNA-protein interactions as well as DNA-DNA interactions within the telomere [81]. Filter binding assays with telomeric oligonucleotides containing deoxyT to riboU substitutions revealed the bolded and underlined nucleotides of GG**<u>TT</u>**AC as most important for RNA discrimination by Pot1Np [81]. Further studies additionally found that mutation of the same two nucleotides in adjacent GGTTAC repeats robustly decreased the binding affinity of full length SpPot1 [82].

Full length SpPot1 was found to stably bind 24nt sequences with ($[GGTTACAC]_3$) and without ($[GGTTAC]_4$) spacer sequences using an electromobility shift assay (EMSA) [61]. In addition to variable spacer sequence, full length SpPot1 also bound sequences ranging from 12 to 24 nts with comparably high binding affinities [83]. However, the interaction between the full length SpPot1 and the 15nt sequence exhibited a 2-fold increase in half-life compared to the 12nt sequence, indicating a more stable interaction between full length SpPot1 and the 15nt sequence. Regarding the nucleotides required for sequence specificity, *Trujillo et al.* indicated that the bolded and underlined nucleotides of G**<u>GT</u>**TACAGGTTACAG were required whereas *Altschuler et al.* suggested a slightly expanded pattern of **<u>GGTTA</u>**C**<u>G</u>**GT**<u>T</u>**AC in the 12nt sequence and G**GTT**ACGGTTACGGT in the 15nt sequence [61, 84]. Nonetheless, both groups concluded that the Pot1 N-terminal domain conferred most of the sequence specificity while the second DNA binding domain accommodated greater sequence diversity and served to extend the protective surface of Pot1. Consistent with this idea, full length Pot1 could bind GGTTCTACA mutant repeats as well as telomere sequences from *T. thermophila, S. cerevisiae,* and *O. nova* while Pot1Np specifically bound GGTTAC [61].

I.3.3 End protection

Natural chromosome ends can be detected as double-strand breaks in need of repair. However, unwarranted repair of the chromosome end leads to genomic instability [85]. In vertebrates, yeast, and ciliates, a complex of protective proteins assembles onto telomeric repeats to cap the chromosome end and attenuate DNA damage response (DDR) pathways [86]. This complex is functionally and structurally conserved in fission yeast and mammals [63]. Taz1 (TRF1 and TRF2 in human) and Pot1 (POT1 in human) are two essential proteins in the capping complex that directly bind telomeric DNA. These proteins protect against distinct DNA double-strand break repair mechanisms.

Classical non-homologous end joining (c-NHEJ) is one mechanism for DNA double-strand break repair. It leads to chromosome end fusions that decrease cell viability in *S. pombe* [87]. In human, fused chromosome ends result in dicentric chromosomes and breakage-fusion-bridge (BFB) cycles that cause further chromosomal aberrations [85]. In *S. pombe*, abundant chromosome end fusions were detected in *taz1* deletion cells arrested in G1 by nitrogen starvation [87, 88]. Chromosome end fusions were shown to be mediated by Ku70 and DNA ligase IV, two factors required for c-NHEJ [87]. Similarly, studies in mice and human have demonstrated that one of the mammalian Taz1 homologs, TRF2 inhibits c-NHEJ-mediated chromosome end fusions [89-91]. TRF1 deletion in mice also causes chromosome end fusions [92]. Human TRF2 additionally contributes to telomere protection by mediating the formation of a t-loop structure where the 3’ overhang invades the double-stranded part of the telomere to evade detection by DDR machinery [93, 94]. Lastly, whether substantial telomere decompaction results from TRF2 depletion and if it is required for DNA damage response signaling at the telomere remains controversial [95-97].

In addition to NHEJ, DNA double-strand breaks can be repaired by homology directed repair mechanisms, such as single strand annealing (SSA), microhomology-mediated end-joining (MMEJ), and homologous recombination (HR). These mechanisms involve the DNA damage sensor, ataxia telangiectasia and Rad3-related protein (ATR), resection of the damaged DNA to form 3' overhangs, and use of the neighboring strand as a template to fill in the double strand break [98]. Additionally, unlike c-NHEJ caused by Taz1/TRF2 telomere uncapping, HR, MMEJ and SSA appear to be triggered by telomere attrition and Pot1 depletion and have the potential to cause further telomere loss. Veritably, *S. pombe pot1* deletion results in rapid telomere shortening and chromosome circularization [71]. Circularization of linear chromosomes involves factors required for SSA but not Ku or ligase 4 proteins required for c-NHEJ [99]. This process results in the loss of telomeric repeats and pairing of microhomology regions (H1-H1', H3-H3', H4-H4', and H5-H5') within subtelomeric DNA at the ends of a single chromosome. In human, critically short telomeres also developed fusions involving microhomology sites [100]. Similarly, the murine orthologs, Pot1a and Pot1b, have been shown to prevent ATR activation and homologous recombination detected as telomere sister-chromatid exchanges [101, 102].

I.3.4 Regulation of telomerase recruitment and activation by *S. pombe* telomeric proteins

The relatively low abundance of telomerase molecules and chromosome ends in the nucleus underscores the need for targeted recruitment, activation, and retention of telomerase specifically at the chromosome end (telomere) [103]. In budding yeast, there are ~30 telomerase holoenzymes and 64 telomeres [104]. In human, despite an excess of telomerase RNA and catalytic subunits, there are only ~240 holoenzymes and 256 telomeres in HEK 293T cells or 304-320 telomeres in HeLa cells [105]. Additionally, telomere length influences telomerase recruitment, albeit differently amongst organisms and various cell types. The telomerases of

budding yeast and murine embryonic stem cells tend to preferentially elongate the shortest telomeres, whereas human telomerase was found to add ~50-60nt to most chromosomes per division in tumor cells [106-110]. In both yeast and human, telomerase recruitment is thought to be mediated by a change in the folding structures of the telomere that either exposes or conceals the 3' overhang [107, 111]. While the molecular details of this process still need to be determined in human, they have been characterized in fission yeast.

Each of the five core telomere capping proteins in *S. pombe* are involved in telomere length maintenance. Individual deletion mutants of *taz1, rap1* and *poz1* exhibit dramatically increased telomere lengths [64, 69, 70, 80]. The increased telomere lengths of *taz1*- and *poz1*- mutants, respectively, were abrogated when the telomerase catalytic subunit (trt1) was simultaneously deleted, implicating both Taz1 and Poz1 in the negative regulation of telomerase activity [62, 80]. Conversely, deletion of *tpz1* resulted in the rapid and complete loss of telomeres giving rise to survivors with circularized chromosomes [80]. However, a minishelterin complex of Taz1 tethered to Tpz1 in the absence of Poz1 maintained wild type telomere length, arguing against a direct requirement for Poz1 in length maintenance [112]. The control of telomerase activity by Pot1 is more complex. Deletion mutants of *pot1* immediately lose telomeres and form circularized chromosomes, suggesting a role in the positive regulation of telomerase activity [71, 99]. In a direct telomerase activity assay with myc9-tagged Trt1, addition of full length Pot1 reduced telomerase activity in a dosage-dependent manner, indicating that the 3' overhang might be obscured from telomerase binding thereby conferring a negative regulatory role to Pot1 [61].

A growing amount of evidence supports a three-phase model for telomerase recruitment, retention, and activation at chromosome ends in *S. pombe* (Figure 1.3) [113]. Firstly, during late

S-phase, telomeres are prepared for telomerase recruitment by the phosphorylation of Ccq1. The Ccq1 protein was previously identified as a binding factor in the yeast spindle pole body and found to co-localize with Taz1 during telomere pairing in meiosis [114]. It also localizes to telomeres through a direct interaction with Tpz1 [115]. In the absence of Ccq1, telomeres shorten and signaling of the G_2 checkpoint kinase, Chk1, is activated [116]. The Ccq1 protein is phosphorylated on the Threonine 93 residue (Thr93) by Tel1, an ortholog of the ATM PI3-kinase, and Rad3, an ortholog of the ATR PI3-kinase [117, 118]. Thr93 phosphorylation is negatively regulated by Taz1, Rap1, and Poz1 through their inhibition of telomeric Rad3 activity [119].

Secondly, telomerase is recruited to the telomere through a simultaneous Ccq1-Est1 and Tpz1-Trt1 interactions [113]. The Est1 protein is evolutionarily conserved with *S. cerevisiae* and human and serves as an essential factor *in vivo* for *S. pombe* telomerase holoenzyme activity [120]. The Ccq1-Est1 interaction is facilitated by Thr93 phosphorylation and is required for Est1 association with the telomere and thereby telomerase activity at the telomere [121]. Tpz1 binds Trt1 through an evolutionarily conserved Tel patch [113].

Thirdly, telomerase is stabilized and activated at the chromosome end [113]. The Ccq1-Est1 interaction is transient and telomerase becomes stabilized at the telomere through its interaction with Tpz1 [122]. Telomerase can then align the RNA template to the 3' overhang, where its activation requires ccq1, Tpz1, and Est1 [122].

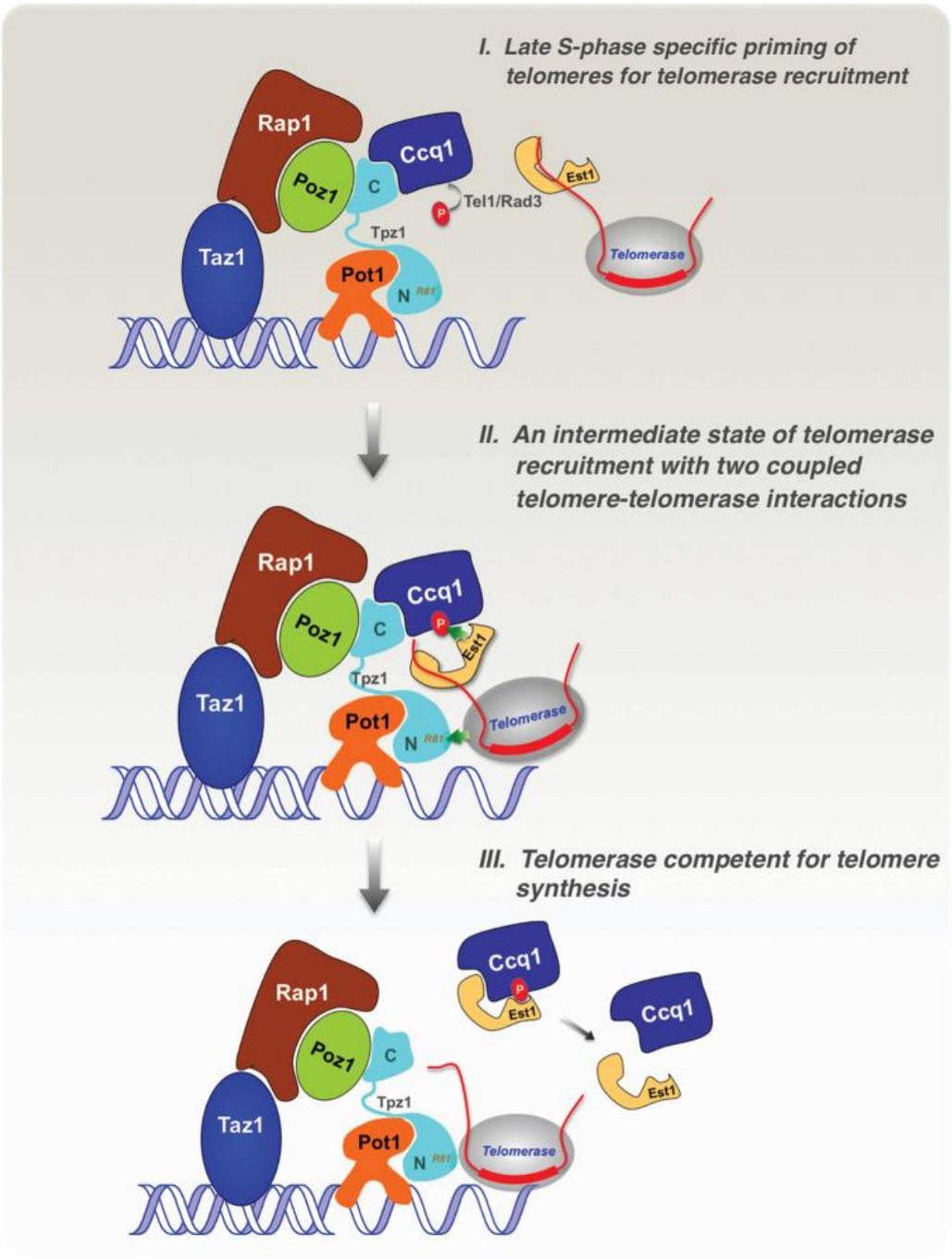


Figure 1.3 Three-state model for the recruitment, retention, and activation of telomerase in *S. pombe*. First, Ccq1 is phosphorylated to facilitate its interaction with the telomerase holoenzyme. Second, Est1 of the holoenzyme interacts with Ccq1 and Trt1 of the holoenzyme interacts with Tpz1. Third, telomerase is stably associated with the 3' overhang through its interaction Tpz1. Figure is adapted from Hu *et al.* [113] under the Creative Commons Attribution License.

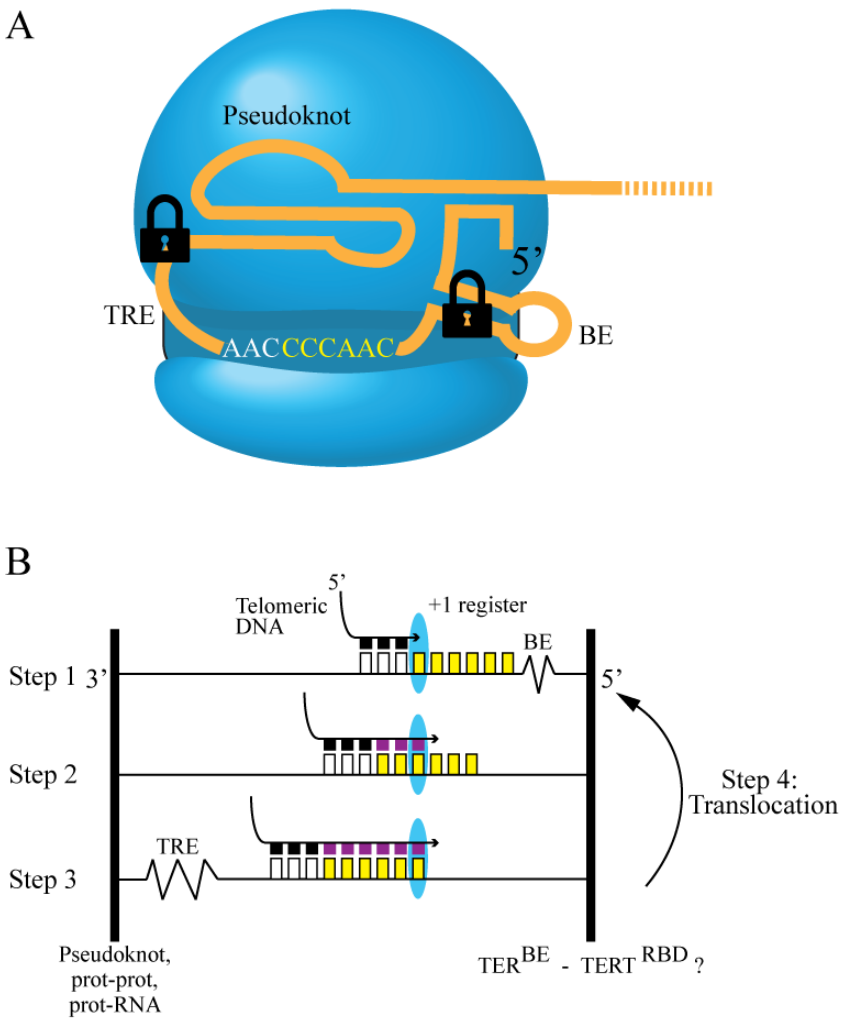
I.4 Telomerase catalysis

All telomerases catalyze the addition of telomeric DNA repeats through the reverse transcription of an internal RNA template. However, the biochemical interactions between a telomeric primer, telomerase RNA, and telomerase catalytic subunit can vary among ciliates, vertebrates, and fungi. These variations affect the repeat addition processivity of different telomerases. Repeat addition processivity (RAP) is the synthesis of multiple copies of a templated sequence onto the 3'end of an oligonucleotide primer *in vitro* or a chromosome end *in vivo* prior to telomerase dissociation [58, 123]. Both *Tetrahymena* and human telomerases have been demonstrated to be processive with *in vitro* biochemical assays [58, 124]. Regarding fungal telomerases, whether these enzymes are processive *in vivo* and what confers processivity *in vitro* are controversial. The red bread mold, *N. crassa*, has been shown to be processive *in vitro [43]*. However, *S. cerevisiae* telomerase has been shown to be nonprocessive *in vitro* as well as generate highly heterogeneous repeats *in vivo* (Chapter I.2.3, [125]). RAP has not been demonstrated for *S. pombe* telomerase *in vitro.* Furthermore, *S. pombe* is also known to generate highly heterogeneous repeat sequences *in vivo* (Chapter I.2.3). Models for telomerase catalytic activity have been proposed for *Tetrahymena*, human, and *S. pombe [40, 46, 123, 126]*.

I.4.1 Accordion model in *Tetrahymena*

The accordion model was first proposed by Berman *et al.* after a series of telomerase activity assays and single-molecule Förster resonance energy transfer (smFRET) assays [123]. The model states that two regions beyond the template, one upstream and one downstream, are fixed structures in the telomerase RNA (Figure 1.4A). In *Tetrahymena*, the upstream fixed structure is a physical boundary element created by a hairpin loop. The downstream fixed structure may be the pseudoknot which directly interacts with the catalytic subunit. The template

sequence between these two fixed structures can undergo extension and compression in the same way that an accordion is used and thereby move the +1 alignment register into the enzyme active site (Figure 1.4B). Just as with a compressed spring, the energy for the translocation step could be provided by the compression of the RNA sequence downstream of the template. As each nucleotide of the template is moved into the active site, the downstream sequence becomes compressed and then when the compressed downstream sequence recoils the template is re-aligned to the 3'overhang. Still, too little is known about the structure of the RNA in complex with the catalytic subunit to make conclusions about the exact conformational changes of the RNA downstream of the template.



(B) is adapted by permission from Macmillan Publishers Ltd: [Nat Struct Mol Biol] Berman, A.J., et al., The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis. Nat Struct Mol Biol, 2011. 18(12): p. 1371-5., copyright (2011)

Figure 1.4 The accordion model for *Tetrahymena* telomerase activity. A) Partial structure of *Tetrahymena thermophila* telomerase. Telomerase catalytic subunit is shown in blue (created by Mark Miller, Stowers Institute for Medical Research).Telomerase RNA is shown in orange. Stem IV and the 3' end are omitted for simplicity. Boundary element (BE) is one anchor point denoted by a lock symbol. A second anchor point may be the pseudoknot. The template recognition element (TRE) is used to align the RNA to the telomeric 3' overhang. The alignment region is the sequence written in white and the template is the sequence written in yellow. B) Alignment denoted by white rectangles. Template denoted by yellow rectangles. Step 1: alignment of RNA to telomeric DNA, compression of BE. Step 2: As template moves through the active site and + register, BE lengthens. Step 3: When the 5' most nucleotide of the template enters the +1 register, the TRE is maximally compressed. Step 4: In a translocation step, the TRE recoils to push the 3' most nucleotide of the template into the +1 register.

I.4.2 Telomerase catalytic cycle in human

Unlike the *Tetrahymena* telomerase, which uses a nearby, physical boundary element, human telomerase uses a pause site inherent in the telomerase RNA (hTR) template sequence to limit the length of an individual repeat [127]. Again with the use of smFRET, researchers have described a model for the catalytic cycle of human telomerase [126]. First, the hTR template aligns to a DNA primer. Second, in a relatively fast step, the complete repeat is synthesized and realigned to the telomerase RNA. Third, the realigned telomeric DNA: telomerase RNA hybrid is shifted in relationship to the catalytic subunit to bring the +1 register into the active site and move the proximal part of the telomere out of the active site. This step was found to be the rate-limiting step in RAP. In addition to the catalytic cycle, the shelterin component TPP1 can also promote RAP independently of its role in end protection [124, 128]. Impaired RAP has been described in Hoyeraal Hreidarsson Syndrome and pulmonary fibrosis [129, 130]. Lastly, human telomerase RAP is also sensitive to treatment with the telomerase inhibitor, BIBR1532 [131, 132].

I.4.3 Synthesis of degenerative repeats by *S. pombe* telomerase

In contrast to the pause site of vertebrate telomerases, a physical boundary element (BE) in the telomerase RNA subunit defines repeat length in *S. pombe.* The boundary element is a hairpin structure upstream and in close proximity to the template. Traditionally, experiments that have identified BEs have introduced mutations that destabilized the paired region or increased its distance from the template resulting in repeats with new sequences and increased lengths. Compensatory mutations that restored the paired region in turn restored repeat length and sequence as well. In this way, BE structures were found to be conserved in species of *Tetrahymena* [133] and *Kluveryomyces* [134] as well as in *S. cerevisiae [135]and S. pombe [46]*.

However, the mechanism for boundary element function is distinct between ciliates and yeast. A study of the interaction between the RNA binding domain of the catalytic subunit and the telomerase RNA revealed that the *Tetrahymena* BE behaves in a sequence-specific manner [136, 137]. Conversely, replacement of the wild type sequence with a variant sequence that maintained a base paired structure has demonstrated that structure rather than sequence governs BE function in *Kluveryomyces lactis* and *S. cerevisiae* [134, 135].

Three key features distinguish wild type *S. pombe* telomerase from repeat addition processive telomerases (Figure 1.5). The alignment region can overlap with the template by a single nucleotide at the 3' end of the template. The BE can adopt two conformations, open and closed, and when in the closed conformation, the BE and template overlap at the 5' end of the template by a single nucleotide. The two conformations of the boundary element exist in an equilibrium that favors the open conformation [46]. The closed conformation confers a 5'-GGTTAC-3' sequence that forms the core of all wild type repeats. The open conformation results in the synthesis of 5'-GGTTACA-3' and 5'-GGTTACAC-3' repeats, accounting for the 5'-AC-3' spacer nucleotides. The 3' end of the newly synthesized repeat dictates the alignment register of the following translocation step thereby affecting the beginning sequence of each repeat. When the repeat ends with an A, which is most frequently the case, the next repeat is more likely to include a variable number of guanines [45, 46, 138]. The variable number of guanines arises through a process termed "stuttering" [40, 46]. Stuttering is when the alignment register is shifted a single nucleotide upstream to accommodate the 3'-A of the telomere. Also, "slippage" occurs when the telomeric 3' overhang translocates from C239 to C240 generating a variable number of consecutive guanines into the telomere. Together the flexible BE, stuttering, and slippage provide the mechanism by which spacer nucleotides are introduced into the

telomere. The frequency of heterogeneous telomeres among fungal species has led to the assumption that the nonprocessive nature of fungal telomerases is an inherent property of the enzyme.

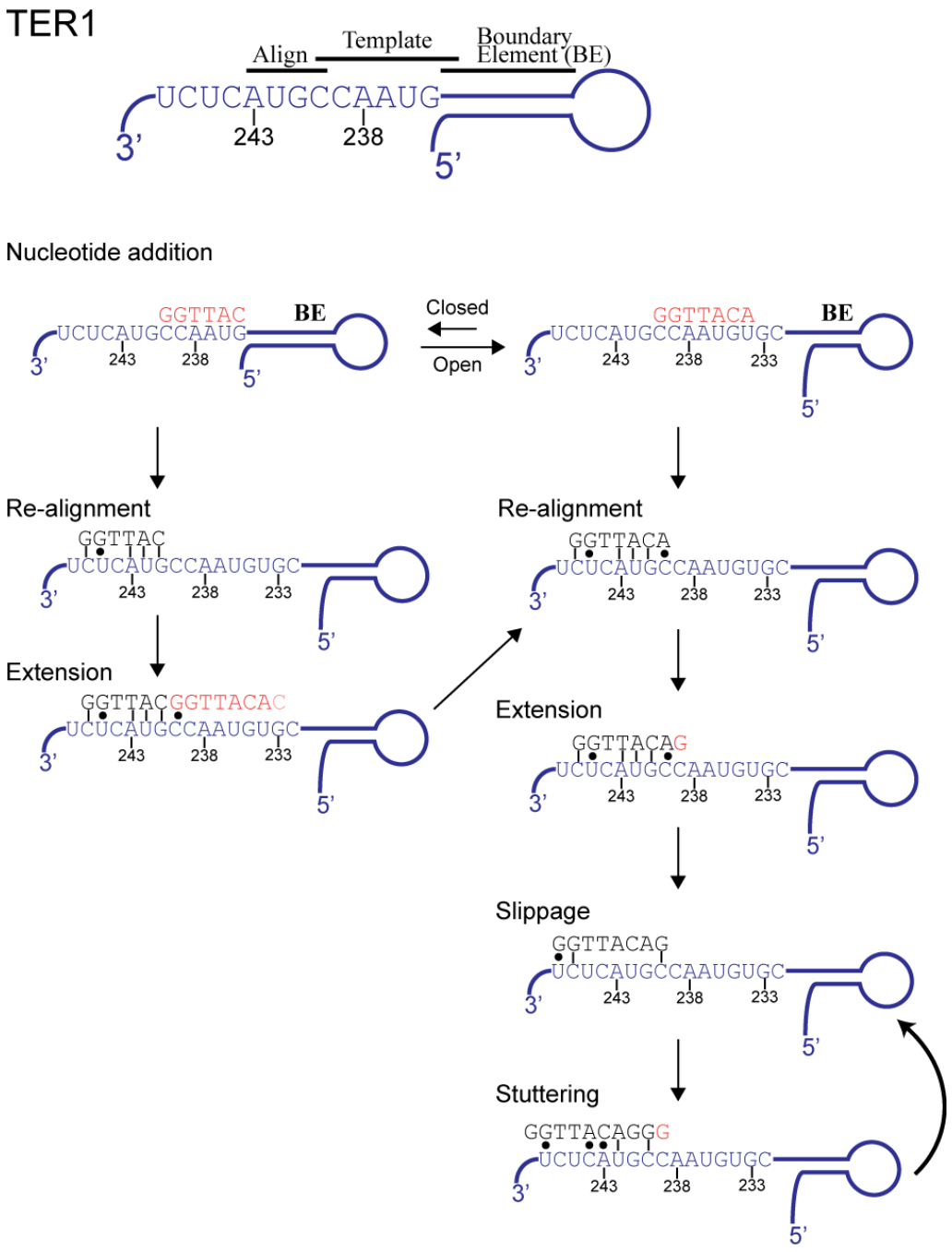


Figure 1.5 Telomerase catalytic activity in *S. pombe*. Telomerase RNA (blue), newly synthesized repeat (orange), and telomeric DNA (black). The telomerase RNA subunit (TER1) has an alignment region, a template, and a physical boundary element (BE). The boundary element is shown as a hairpin with a single loop for simplicity. The boundary element exists in an equilibrium that favors the open conformation over the closed conformation. The closed conformation results in a 5'-GGTTAC-3' repeat whereas the open conformation can result in 5'-GGTTACA-3' or 5'-GGTTACAC-3'. If the last nucleotide of the repeat is an "A", then the enzyme undergoes stuttering and slippage to add 2-6 guanines before the next repeat.

I.5 Telomeres and telomerase in cancer

The many roles of telomeres and telomerase in fundamental cellular processes implicate them in a range of human diseases. At one end of the spectrum are premature aging syndromes, such as Dyskeratosis Congenita and Hoyeraal Hreidarsson Syndrome, and telomeropathies, including aplastic anemia, idiopathic pulmonary fibrosis, and liver cirrhosis. These diseases are caused by insufficient telomerase activity in stem cell populations required for organ growth and repair [139]. Consequently, seemingly unrelated clinical signs and symptoms tend to manifest in high turnover tissues, such as the bone marrow, skin, and gastrointestinal tract, as well as tissues affected by chemical injury, such as the liver and lungs. At the other end of the spectrum are 85-90% of cancers that aberrantly increase telomerase activity [140, 141]. Therefore, the timing of telomerase expression and activity must be closely regulated for normal cell and, ultimately tissue, function.

I.5.1 The role of telomeres and telomerase in malignant transformation

Growing evidence suggests that telomeres play two somewhat paradoxical roles in tumorigenesis, the key to which depends on the mutational background of the tumor (Figure 1.6) [85]. In normal somatic tissue, prior to the loss of RB and p53 expression and the reactivation of telomerase catalytic subunit (hTERT) expression, short telomeres serve as tumor suppressors. As the telomeres shorten they become uncapped by protective proteins, allowing DNA Damage Response (DDR) effectors to access the chromosome end and signal cell cycle arrest [142-144]. This event results in the senescence of cells with short telomeres thereby preventing cellular proliferation. However, in the absence of pRB or TP53 function, cells harboring short telomeres can continue to divide, bypassing cell cycle checkpoints.

The continued division in the presence of unprotected chromosome ends results in genomic instability which is thought to generate mutations important for malignant transformation [85]. More specifically, homology directed repair at the chromosome end through mechanisms such as SSA or MMEJ, generate fusions between uncapped chromosome ends to produce dicentric chromosomes incorporating deletions and insertions around the fusion point [99, 100]. Furthermore, the resolution of these dicentric chromosomes has been associated with genomic changes that typify malignant tumors including aneuploidy and loss of heterozygosity [145, 146], chromosomal rearrangements [147], chromothripsis (when one or more chromosomal regions becomes fragmented and then haphazardly stitched together during repair, [148]), kataegis (hypermutation of TpC dinucleotides to C>T or C>G, [148]), and amplifications and deletions [149]. Maciejowski and de Lange propose that activating telomerase mutations may be a product of the genomic instability experienced during telomere crisis as well as an exit strategy from telomere crisis [85]. As telomerase heals the chromosome end, cells would emerge from telomere crisis with highly mutated genomes that are fertile ground for selection and tumor progression. Although the precise timing of telomerase reactivation has yet to be established, the detection of *TERT* promoter mutations in premalignant lesions and early-stage malignancy support this concept [150-152].

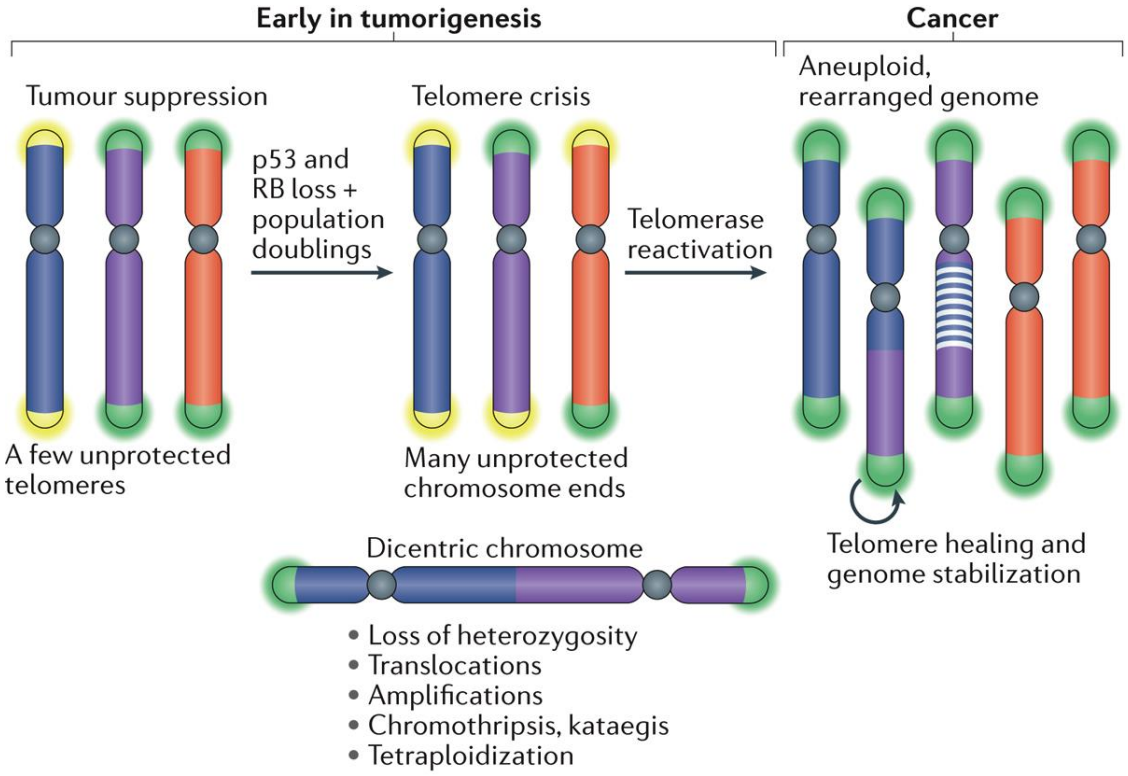


Figure 1.6 The dual roles of telomeres in tumor progression. Early in tumorigenesis, telomeres serve a protective role by limiting replicative potential. As cells acquire loss of function mutations in p53 and RB, cells bypass cell cycle checkpoints and continue to divide despite critically short telomeres. During telomere crisis, telomeres develop end fusions and instigate genomic instability, one hallmark of cancer. This genomic instability is hypothesized to lead to telomerase reactivation and evasion of replicative senescence, a second hallmark of cancer.

I.5.2 Telomerase reactivation in cancer

Reactivation of telomerase activity is achieved primarily through the upregulation of hTERT transcription and several mechanisms underlie its aberrant transcription in various cancers. Genomic rearrangements, some of which position a super-enhancer within 1 megabase upstream of the gene, cause hTERT overexpression in neuroblastoma [153, 154]. Additionally, hTERT transcription in cancer can be activated exogenously through proteins encoded by

oncogenic viruses including human T-cell lymphotropic virus type-I (HTLV-I) [155], human cytomegalovirus (HCMV) [156], and human papilloma virus (HPV) [157] [158]. The Wnt/β-catenin pathway also upregulates hTERT through a direct interaction between β-catenin and the *TERT* promoter [159, 160]. c-Myc appears to exert opposing effects on *TERT* promoter activity. A study of tumor cell lines demonstrated that a histone methyltransferase called the SET and MYND domain-containing protein 3 (SMYD3) was required for the maintenance of the Histone 3-Lysine 4 trimethyl (H3K4me3) mark of active chromatin which in turn recruited c-MYC to the *TERT* promoter and increased mRNA levels [161]. When in a complex with the myc-associated factor X (Max), c-Myc could either activate or repress the *TERT* promoter. However, when in a complex with the breast cancer susceptibility protein 1 (BRCA1), c-Myc represses *TERT* promoter activity [162, 163].

Perhaps the most, well-characterized mechanism for hTERT reactivation is the epigenetic switch conferred by point mutations in the promoter. Normally, hTERT is silenced in somatic tissue through the histone H3K27me3 mark [164]. Two seminal papers identified two mutations, -124 C>T and -146C>T, in the *TERT* promoter through whole-genome sequencing of malignant melanomas and a linkage study of a family of melanoma patients [165, 166]. Since then, mutations in the *TERT* promoter have become the most frequent known noncoding mutations in cancer [167-169]. The originally identified, -124 C>T and -146 C>T, are also the most common mutations and have been shown to be sufficient for immortality in human pluripotent stem cells and increased mRNA expression in human cancer [170-172]. These mutations lead to a SMYD3-mediated switch from the silent mark, H3K27me3, to the active mark, H3K4me2/3, and create a binding site for the E twenty-six (ETS) transcription factor complex, GABPA/B1 [161, 164-166, 173]. Transcription factor association in turn results in pol II recruitment and

monoallelic hTERT expression [164, 174]. Still, questions regarding how to therapeutically target telomerase promoter mutations and if this would be an effective therapeutic strategy remain unanswered [169].

I.5.3 The role of telomeres and telomerase in tumor progression

Research has focused on the contribution of telomeres and telomerase to two hallmarks of cancer, evasion of replicative senescence and genomic instability. Yet, telomeres and telomerase play a role in many other aspects of cancer as well. Single stranded, G-rich telomeric DNA is particularly susceptible to oxidative damage [175-177]. When an oxidized nucleotide, 8-oxo-2'deoxyguanosine-5'-triphosphate (8oxodGTP), is incorporated into the telomere, telomerase reverse transcription is stalled [175, 178]. However, if a guanine of the telomere is directly oxidized, then telomerase activity is increased, presumably due to the destabilization of inhibitory telomeric, G-quadruplexes [178]. hTERT is protective against reactive oxygen species (ROS) and ROS-induced apoptosis [179-181]. Additionally, hTERT was shown to abrogate apoptosis in a fashion independent of its telomerase activity in cancer cell lines and mouse embryonic fibroblasts [182-184].

Inflammation causes telomere shortening and is an important factor in the tumor microenvironment [185, 186]. More specifically, pro-inflammatory cytokines like IL-6 have been reported to activate telomerase activity as measured by the telomeric repeat amplification protocol (TRAP) assay [187]. Rap1, the telomere-associated protein that forms the double stranded- telomeric DNA sub-complex with TRF2 in humans, is involved in positive reciprocal interactions with mediators of the NF-κB pathway and is associated with malignant grade breast cancer [188]. A study of HEK293T cells and HeLa cells showed that hTERT directly binds the p65 subunit of NF-κB and augments the expression of cytokines, like interleukin-6 and 8 (IL-6,

IL-8) and tumor necrosis factor α (TNFα) [189]. NF-κB can increase murine TERT (mTERT) expression as demonstrated in mice, suggestive of a positive feedback loop [190].

hTERT also contributes to metastatic disease [191]. It serves as both a transcriptional regulator and target of Wnt-β-catenin signaling [159, 160, 192]. hTERT localizes to Wnt-dependent genes including *matrix metalloproteinase 7* (*MMP-7*), *vascular endothelial growth factor* (*VEGF*), and *epidermal growth factor receptor* (*EGFR*) to facilitate invasion, angiogenesis, and proliferation [193, 194]. It also stimulates MMP expression through its interaction with NF-κB [189] and promotes epithelial-mesenchymal transition (EMT) [195, 196]. Furthermore, *TERT* promoter mutations were found to be more prevalent among patients with invasive urogenital cancer and advanced tumor stages [197] and among metastasized melanomas than primary tumors [168]. Lastly, ribozyme-suppression of telomerase activity resulted in decreased invasive phenotype of the B16 murine melanoma cell line [198]. However, the precise causal and temporal relationship between telomere maintenance and metastasis has yet to be determined [191].

I.5.4 Therapeutic strategies

Although several anti-telomerase strategies have been explored [199], none have been established as mainstays in oncology treatment. RNA interference is one of the few methods that can directly counteract the upregulation of *TERT* gene expression. However, it is likely too unstable to generate sustainable anti-tumor activity. If delivery methods can be optimized, gene therapy with oncolytic adenoviruses under the control of the *TERT* promoter may prove to be a viable option to selectively kill cancer cells. Immunotherapy against the catalytic subunit (e.g. GV1001) is an ideal approach because it could eradicate the reverse transcriptase and non-reverse transcriptase, oncogenic effects of hTERT. However, hTERT-targeted vaccines have

proven to be only marginally beneficial in clinical trials [200]. Imetelstat/GRN163L is an antisense oligonucleotide against the hTR template and is listed as an interventional drug in several, ongoing clinical trials [201]. G-Quadruplex stabilizers and small molecule inhibitors could prevent telomerase-dependent telomere elongation by targeting the holoenzyme. Yet, screening assays to identify such compounds have been largely unproductive due to an inability to detect short, newly synthesized 3' overhangs amidst preexisting kilobase-long telomeric repeats.

In addition to telomerase and its components, telomeres themselves also present a potential therapeutic target. Methods to destabilize telomeres could exploit the natural tumor-suppressive features of unprotected chromosome ends. One type of destabilizer could be a mutant-template hTR that is used to incorporate mutant telomeric repeats and impair sequence-specific shelterin binding. If developed in conjunction with telomerase inhibitors and p53 gene therapy, this therapeutic cocktail could target telomerase positive cancer cells. However, translation of this combinatorial therapy into the clinic is still a far-reaching goal. Two classes of drugs aimed at increasing p53 expression and activity are listed in the NCI Drug Dictionary database: 1) p53-human double minute 2 (HDM2) interaction inhibitors which prevent proteasome-mediated degradation of mutant p53 and 2) p53 re-activation and induction of massive apoptosis (PRIMA-1) analogue which restores the wild type conformation to mutant p53 [202]. Of these two classes, only the PRIMA-1 analogue is currently being tested in a clinical trial (NCT02098343 as of April 26, 2017; www.ClinicalTrials.gov). Additionally, clinical trial-ready telomere destabilizers have not yet been developed. A better understanding of how the telomeric sequence affects telomere uncapping and subsequent activation of DDR, double strand

break repair, senescence, and apoptotic programs would inform the development of telomere destabilizing agents.

I.6 Scope of dissertation

The central goal of my dissertation research is to define how specific sequence patterns of the telomerase RNA template affect telomerase activity and telomere function. This study has yielded results that inform our models of telomere evolution and telomerase catalytic activity in *S. pombe* as well as provided insight into telomere uncapping strategies for cancer treatment. Chapter I provides a review of the current understanding of telomere and telomerase concepts pertinent to this study. Chapters II, III, and IV describe and explain our key findings from multiple telomerase RNA template competition experiments. Each of these chapters includes a materials and methods section, which details the design of major experiments. Lastly, Chapter V summarizes the major contributions of this work to the telomere and telomerase field and proposes several future directions to address questions that emerged from this study.

Chapter II: *In vivo* selection identifies alternative, functional telomerase RNA templates in *Schizosaccharomyces pombe*

II.1 Abstract

Most eukaryotic telomeres are guanine-rich (G-rich), repetitive, and bound by protective and regulatory proteins. Telomerase carries an internal RNA (TER1) template to reverse transcribe DNA repeats and is the most wide-spread mechanism for elongating telomeres. Conservation of these features in telomere biology underscores their functional significance. Which of these features exerts the greatest selective pressure and whether a G-rich telomere is important for competitive growth are important aspects of telomere evolution that have yet to be studied. An *in flask* evolution experiment in *S. pombe* was used to probe the diversity of telomere sequences and corresponding templates that could maintain competitive growth. Here we report that C-rich templates are enriched among competitive strains. The 3'-CCAAU-5' in the 7 nucleotide template conferred the greatest growth advantage in rich and minimal media. Beyond this core, the position of two variable nucleotides provided a more subtle difference in growth and was associated with telomerase alignment to the chromosome end. Lastly, we provide, for the first time, evidence that the *S. pombe* alignment region can shift as many as four nucleotides to promote reverse transcription of a new template sequence.

II.2 Introduction

The most common telomeric repeat sequence is GGTTAG, present in all vertebrates, and some invertebrates, fungi, and plant species. This predominance suggests that selective pressures drove many systems to use this sequence for end replication and end protection. The enzymatic cycle of telomerase activity, the formation of regulatory folding structures such as G-quadruplexes, and the binding specificity of protective and regulatory proteins each impose constraints on the diversity of sequences found at natural chromosome ends. Yet, despite the apparent success of the GGTTAG repeat, different sequence repeats are found in diverse genera of algae, fungi, ciliates, invertebrates, and plants. Perhaps the most substantial repeat diversity is found among fungi ranging from perfect 25 nucleotide repeats in *Kluveryomyces lactis* to highly heterogeneous repeats in many ascomycetes including the model species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* [59]. Whereas perfect repeat sequences result from faithful reverse transcription of the telomerase RNA template sequence, our lab previously demonstrated that the synthesis of heterogeneous repeats requires stuttering, slippage, and poorly defined boundary elements in the telomerase RNA (Chapter I.4). Additionally, the single-stranded, telomeric DNA binding protein, Pot1, is structurally and functionally conserved between *S. pombe* and human. However, *S. pombe* Pot1 employs different binding modes to accommodate degenerative repeats of varied sequence, length, and number. *S. pombe* is an ideal model system to study the factors that influence telomere sequence and to define the corresponding features of the telomerase RNA template that support competitive growth.

To investigate these factors, we employed the fast growth and responsiveness to telomere-dysfunction of *S. pombe* in an *in flask* evolution experiment. This experiment starts with a population of cells expressing different telomerase RNA template sequences and grows

them competitively in liquid culture. The fastest growing cells must maintain functional telomeres that are elongated and capped. Because these cells continue to divide without entering telomere crisis, the templates contained in them will become the most abundant and the winners of the competition. In this chapter, we determined the fitness of different patterns within the template under rich media and minimal media growth conditions. We also identify two competitively growing clones, whose telomere sequences are not predicted to bind Taz1 or Pot1, and discuss a possible explanation for their success in the competition.

II.3 Materials and Methods

II.3.1 Plasmid Library Construction and Yeast Strains

Plasmid *ter1*$^{+}$ template libraries were generated by cloning. To minimize the risk of wild type overrepresentation in the library after an incomplete plasmid digestion, an intermediate *ter1*$^{+}$ plasmid with a deletion of nucleotides 230 to 280, including the template region, was generated by cloning annealed oligonucleotides BLoli 4350 and BLoli 4351 (Table 2.1) into the wild type *ter1*$^{+}$ plasmid, pJW10 [40]. The resulting plasmid, pMP01, was digested with SfuI and BclI (New England Biolabs) and ligated with a partial oligonucleotide duplex of BLoli 4361, BLoli 4362, and BLoli 4363 (Table 2.1) corresponding to positions 214 to 302 of the *ter1* gene and including randomized nucleotides at positions 234 to 240. This construct was cloned in XL10-Gold Ultracompetent Cells (Agilent Technologies) following the manufacturer recommended protocol. Cloning reactions were set up each with 1:1000 dilution of the annealed oligonucleotide insert and 62ng of digested pMP01. 200µl of cloning reaction was added onto each of 27 LB plates plus 50µg /ml Carbenicillin and incubated overnight at 37°C. Each plate yielded approximately 3,300 colonies. Each plate was washed with 12ml liquid LB plus 50µg /ml Carbenicillin and a cell spreader. Two plate washes were collected into a single 50ml tube and spun at 6,000g for 15min. Media was removed and plasmid DNA was extracted using the

Qiagen Plasmid Maxi kit. Maxipreps were pooled and aliquoted into pMP02 libraries. One aliquot was electroporated into *ter1*$^{-}$ haploid cells sporulated from PP407 [46] for the minimal media competition experiment.

To generate a plasmid library for propagation in rich media, the nourseothricin-MX6 cassette was sub-cloned in place of the *ura4* gene on pMP01. The resulting plasmid, pMP03, was used to prepare the pMP04 library aliquots, A through EE, with the same method used to clone pMP02 libraries from pMP01. The pMP04-A aliquot was electroporated into *ter1*$^{-}$ haploid cells sporulated from PP407 for the rich media competition experiment.

The 48 template library consisting of all seven nucleotide sequences containing 5'-TAACC-3' was created by pooling 1μg of each plasmid, pMP76-pMP123. Each plasmid was cloned individually and sequence verified. To obtain a wild type template plasmid, three oligonucleotides, BLoli 4362, BLoli 4363, and BLoli 5715 (Table 2.1), were annealed to generate an insert of *ter1* positions 214 to 302 with randomized nucleotides at positions 234 and 235. The annealed complex was cloned into pMP03 and bacterial clones were screened for the wild type template by 96-well plasmid prep and Sanger sequencing. Positions 7101 to 1440 of the resulting construct, pMP28, correspond to *ter1* genomic DNA positions -907 to +1440 and were sub-cloned into pTKK21 [99]. The BstBI (2477) restriction site was removed in the resulting plasmid, pMP74, using the Quickchange Lightning site-directed mutagenesis kit (Agilent Technologies) to generate pMP75. The plasmid was verified by BstBI digestion and marker selection. Three oligonucleotides were annealed, BLoli 4362, BLoli 4363, and BLoli 5715 or BLoli 5716, or BLoli 5717 (Table 2.1), corresponding to *ter1* positions 214-302 with randomized nucleotides at positions 234 and 235, 234 and 240, and 239 and 240, respectively. Annealed oligonucleotides were cloned into pMP75 and bacterial clones were screened for all 48

possibilities of TAACC-containing *ter1* templates to generate plasmids pMP76 - pMP123. The pooled pMP76-pMP123 were electroporated into *ter1*$^{-}$ haploid cells sporulated from PP433 [203].

Table 2.1 Oligonucleotides used to generate plasmid libraries

Name	Sequence
BLoli 4350	5’- phosphate- CGAATTCCTGTACTGCAGCTTTTTTTAGAGTTTT
BLoli 4351	5’-phosphate- GATCAAAACTCTAAAAAAAGCTGCAGTACAGGAATT
BLoli 4361	5’ – phosphate - CGAATTCCTGTACTGCTTCGNNNNNNNGTACTCTTCAACTTTTCAGCATTGCGAAATTATTCTTTTTAGCTTTTTTTTAGAGTTTT
BLoli 4362	5’-phosphate- CGAAGCAGTACAGGAATT
BLoli 4363	5’-phosphate- GATCAAAACTCTAAAAAAAGCTAAAAAGAATAATTTCGCAATGCTGAAAAGTTGAAGAGTAC
BLoli 5715	5’-phosphate- CGAATTCCTGTACTGCTTCGNNTAACCGTACTCTTCAACTTTTCAGCATTGCGAAATTATTCTTTTTAGCTTTTTTTTAGAGTTTT
BLoli 5716	5’-phosphate - CGAATTCCTGTACTGCTTCGNTAACCNGTACTCTTCAACTTTTCAGCATTGCGAAATTATTCTTTTTAGCTTTTTTTTAGAGTTTT
BLoli 5717	5’-phosphate - CGAATTCCTGTACTGCTTCGTAACCNNGTACTCTTCAACTTTTCAGCATTGCGAAATTATTCTTTTTAGCTTTTTTTTAGAGTTTT

Table 2.2 *S. pombe* strains used in this study

Name	Genotype	Source	Purpose in this study
PP138	*h⁻ ade6-M216 leu1-32 ura4-D18 his3-D1*	Lab Stock	Wild type control in southern blot for telomere lengths
PP407	*h+/h- leu1-32/leu1-32 ura4-D18/ura4-D18 his3-D1/his3-D1 ade6-M210/ade6-M216 ter1⁺/ter1::kanMX6*	[46]	pMP02 library transformation and pMP04 library transformation
PP433	*h+/h- leu1-32/leu1-32 ura4-D18/ura4-D18 his3-D1/his3-D1 ade6-M210/ade6-M216 ter1⁺/ter1::ura4*	[203]	pMP76-123 library transformation

II.3.2 Culture Media

Diploid strains were sporulated on ME plates (3% w/v malt extract [Sunrise Sci], 2% w/v Bacto agar [BD]). PP407 *ter1⁻* haploid cells were selected on Yeast Extract low Adenine (YEA) plates (0.5% w/v Bacto yeast extract, 3% w/v dextrose, 2% w/v Bacto agar, 225mg/L histidine HCl, 225mg/L leucine, 225mg/L uracil, 74.0μM adenine, deionized water) plus 100μg/ml Geneticin (US Biologicals) plates. PP433 *ter1⁻* spores were selected on Pombe Minimal Glutamate (PMG) plates lacking uracil (14.7mM potassium hydrogen phthalate, 15.5mM $Na_2HPO_4 \bullet 7H_2O$, 2% w/v dextrose, 13.0mM L-glutamic acid monosodium salt hydrate, 5.2mM $MgCl_2 \bullet 6H_2O$, 99.8μM $CaCl_2 \bullet 2H_2O$, 13.4mM KCl, 0.282mM Na_2SO_4, 4.20μM pantothenic acid, 81.2μM nicotinic acid, 55.5μM inositol, 40.8nM biotin, 8.09μM boric acid, 2.37μM $MnSO_4$, 1.39μM $ZnSO_4 \bullet 7H_2O$, 0.740μM $FeCl_2 \bullet 6H_2O$, 0.247μM molybdic acid, 0.602μM KI, 0.160μM $CuSO_4 \bullet 5H_2O$, 4.76μM citric acid, 0.555mM adenine pH < 10.0, 0.716mM histidine, 1.14mM leucine, mineralized water [Stowers Institute for Medical Research, 1000 E 50th St., Kansas City, MO 64110]).

Strains transformed with the pMP02 plasmid library were propagated in Edinburgh Minimal Media (EMM) lacking uracil (14.7mM potassium hydrogen phthalate, 15.5mM $Na_2HPO_4 \bullet 7H_2O$, 2% w/v dextrose, 93.5mM NH_4Cl, H_2O [Ozarka, Nestle Waters North America Inc. 900 Long Ridge Rd, Stamford, CT. 06902], and salts, vitamins, minerals, and supplements as described for PMG plates lacking uracil). Strains transformed with the pMP04 library were grown in Yeast Extract with Supplements (YES) (0.5% w/v Difco yeast extract [VWR/BD], 3.0% w/v dextrose [VWR/EM], 225mg/L histidine HCl, 225mg/L leucine, 225mg/L uracil, 225mg/L adenine, distilled water) plus 50μg/ml ClonNat (Werner). Yeast clones from this competition time course were streaked from glycerol stocks onto YEA plus 100μg/ml ClonNat plates. Strains transformed with the pMP76 – pMP123 library were grown in YES plus 20μg/ml Geneticin.

II.3.3 Monitored Growth in Liquid Culture

All competition experiments were carried out in 100 ml liquid cultures incubated at 32°C with rigorous shaking unless otherwise noted. For the minimal media time course, a *ter1*$^-$ haploid colony was patched onto a YEA Geneticin plate and used to inoculate 200 ml YES. The culture was grown to 5 x 10^6 cells/ml at which time 100ml was used for transformation with the pMP02 plasmid library and the remaining 100ml was grown to stationary phase (1 x 10^8 cells/ml), aliquoted into 20 ml, collected by centrifugation, washed once with water, and stored at -80°C for later genomic DNA preparation of samples prior to transformation. The transformed culture was inoculated into 20 ml EMM lacking uracil and grown for 90 hours (Round 1). Cultures were then counted using a hemocytometer and used to inoculate the next round of fresh selective media at a cell density of 5 x 10^5 cells/ml. Cultures were then grown for 48 hours (Round 2) after which cell density was determined by counting and cells were diluted to

a density of 5 x 10^5 cells/ml into fresh selective media. The remainder of the culture was divided into 1 ml used to prepare glycerol stocks (YES, 10% glycerol v/v) and 20 ml aliquots that were collected by centrifugation, washed once with water, and stored at -80°C for later genomic DNA preparation. These steps beginning with Round 2 were repeated every 48 hours for 66 days.

For the pMP04 library rich media time course, a *ter1*$^-$ haploid colony was patched onto a YEA Geneticin plate and used to inoculate 200ml YES. The culture was counted at 5 x 10^6 cells/ml and 100ml was collected by centrifugation, washed once with water, and stored at -80°C for later genomic DNA preparation of the culture prior to transformation. Strains transformed with the pMP04 library were recovered in YES for six hours at room temperature and shaking at 225rpm. Recovered cultures were counted using the hemocytometer and used to inoculate the first round of selective media at 5 x 10^5 cells/ml. These cultures were grown for 49.5 hours (Round 1) and then counted and used to inoculate fresh selective media at a cell density of 5 x 10^5 cells/ml. Cultures were grown for 24 hours (Round 2) then counted and diluted into fresh selective media at a density of 5 x 10^5 cells/ml. The remaining culture was aliquoted into 1 ml for glycerol stock preparation and 20 ml for later preparation of genomic DNA. The steps starting at Round 2 were repeated every 24 hours for 63 days.

Lastly, for the pMP76 – pMP123 library rich media time course, *ter1*$^-$ spores were selected on PMG lacking uracil then a single colony was patched to YEA. The patch was used to inoculate a 100ml YES culture that was transformed when it reached 0.75 x 10^7 cells/ml. Strains transformed with the pMP76 – pMP123 library were recovered for six hours at room temperature and shaking at 225rpm. Cultures were then counted and diluted to 5 x 10^5 cells/ml in YES Geneticin and grown for 73 hours (Round 1). Cultures were counted and diluted to 5 x

10^5 cells/ml in fresh selective media for the first 9 rounds until they consistently reached ~1.2 x 10^8 cells/ml in 24 hours. Then for subsequent rounds, cultures were diluted 240x in fresh, selective media. The remaining culture was used to prepare glycerol stocks and 20-25ml was pelleted and frozen for later genomic DNA preparation. These steps were repeated every 24 hours for 63 days.

II.3.4 Illumina Library Preparation and Sequencing

Illumina libraries for the pMP02 minimal media competition and the pMP04 rich media competition were generated using the same PCR protocol. PCR reactions (50µl) consisted of 5x Phusion HF Buffer (Thermo Scientific), 200µM dNTPs, 0.5µM each of primers BLoli 4667 and BLoli 4672 (Table 2.2), 0.02 U/µl Phusion Hot Start II High Fidelity DNA polymerase (Thermo Scientific), and 1µl of a 1:50 dilution of ~30-80µg/ml extracted DNA. The BLoli 4667 primer included a 10 nucleotide barcode sequence later used to monitor PCR amplification bias. The cycling parameters were 98°C for 30 seconds, then 15 cycles of 98°C for 30 seconds, 57°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 10 minutes. DNA products were cleaned up with a Qiagen MinElute PCR purification kit. Samples were then diluted 10x and 1µl was added to a second PCR reaction (50µl) of 5x Phusion HF Buffer, 200µM dNTPs, 0.5µM each of primer BLoli 4666 and respectively indexed primer BLoli 4668-4671, 4770, 4774-4792 (Table 2.2), and 0.02 U/µl Phusion Hot Start II High Fidelity DNA polymerase. The cycling parameters were as follows: 98°C for 30 seconds, 28 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for ten minutes. DNA products were gel purified using a Qiagen MinElute Gel Extraction kit, analyzed for quality using the Agilent 2100 Bioanalyzer, and quantified using the Qubit 3.1 Fluorometer (Invitrogen). Custom PCR libraries were combined with other custom libraries of

the same competition time course in a 10 nM pool. The pMP02 minimal media libraries were run with 30% PhiX DNA on Illumina HiSeq 2500 for 50bp single reads. The pMP04 rich media libraries were spiked at 60% into complex genomic libraries from *S. pombe* making up the remaining 40% and run on Illumina HiSeq 2500 for 50bp single reads.

Illumina libraries for the pMP76 – pMP123 rich media competition were synthesized using a modified PCR protocol. PCR reactions (50µl) included 5x Phusion HF Buffer (Thermo Scientific), 200µM dNTPs, 0.5µM each of primers BLoli 4667 and BLoli 4672 (Table 2.2), 0.02 U/µl Phusion Hot Start II High Fidelity DNA polymerase (Thermo Scientific), and 100ng of genomic DNA. The cycling parameters were as follows: 98°C for 1 minute, 68°C for 30 seconds, 72°C for 30 seconds, with a final extension at 72°C for 10 minutes. DNA products were cleaned up with a Qiagen MinElute PCR purification kit. The entire volume of purified products was added to a second PCR reaction (50µl). Reactions included 5x Phusion HF Buffer, 200µM dNTPs, 0.5µM each of primer BLoli 4666 and respectively indexed primer (Table 2.2), and 0.014U/µl Phusion Hot Start II High Fidelity DNA polymerase. The cycling parameters were as follows: 98°C for 30 seconds, 28 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for ten minutes. DNA products were gel purified using a Qiagen MinElute Gel Extraction kit and analyzed for quality by the 4200 TapeStation and D1000 Reagents (Agilent Technologies). Libraries were quantified with Qubit, combined into a 10 nM pool, and then loaded onto the Illumina HiSeq 2500. Libraries were loaded at 20% with 80% from complex RNA-sequencing libraries from *S. pombe*. Libraries were sequenced using 100bp single reads.

Table 2.3 Oligonucleotides used to prepare Illumina PCR libraries

Name	Sequence
BLoli 4666	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
BLoli 4667	5'- CTACACGACGCTCTTCCGATCTNNNNNNNNNNCGAATTCCTGTACTGCTTCG
BLoli 4672	5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGGCAGAAGACTCACGT
BLoli 4668	5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTG
BLoli 4669	5'-CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTG
BLoli 4670	5'-CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTG
BLoli 4671	5'-CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTG
BLoli 4770	5'-CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTG
BLoli 4774	5'-CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTG
BLoli 4775	5'-CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTG
BLoli 4776	5'-CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTG

BLoli 4777	5’-CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTG
BLoli 4778	5’-CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTG
BLoli 4779	5’-CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTG
BLoli 4780	5’-CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTG
BLoli 4781	5’-CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTG
BLoli 4782	5’-CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGTG
BLoli 4783	5’-CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGACGTGTG
BLoli 4784	5’-CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGACGTGTG
BLoli 4785	5’-CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTG
BLoli 4786	5’-CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGACGTGTG
BLoli 4787	5’-CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGACGTGTG

BLoli 4788	5'-CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTG
BLoli 4789	5'-CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTG
BLoli 4790	5'-CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTG
BLoli 4791	5'-CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCAGACGTGTG
BLoli 4792	5'-CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTCAGACGTGTG

* The underline denotes the index sequence used for multiplexing. The double underline denotes the barcode sequence used to identify preferentially amplified templates after PCR.

II.3.5 Computational Analysis

Raw data were de-multiplexed into individual libraries for each round and converted to FASTQ files. Variant seven nucleotide templates were located by mapping reads to the *Schizosaccharomyces pombe* reference genome ASM294v2 and verifying nucleotides at positions 214 to 233 of *ter1* upstream of the template (upstream flank) and nucleotides 241 to 254 of *ter1* downstream of the template (downstream flank). The sequence of the randomized barcode was determined by locating the 10 nucleotides upstream of the invariant upstream flanking sequence. Raw reads were filtered in 4 steps for the pMP04 rich media (Figure 2.1) and pMP02 minimal media (Figure 2.2) libraries. First, all reads were required to have a 10 nucleotide barcode, a 20 nucleotide upstream flank, a 7 nucleotide template, and a 14 nucleotide downstream flank. Zero mismatches were allowed in the flanking sequences. Second, all 7

nucleotides of the template were stipulated to have a phred quality score of ≥ 30 which corresponds to a 99.9% base call accuracy [204, 205]. Third, to eliminate potential PCR bias only unique barcode-template pairs were used. Multiple reads with the same template-barcode pair were counted only once. Fourth, to minimize noise from reads near the detection limit, a minimum of 50 reads was required of templates in the minimal media experiment and a minimum of 64 reads was required of templates in the rich media experiment. Steps 1-3 were used to filter raw reads from the pMP76-pMP123 rich media library (Figure 2.3).

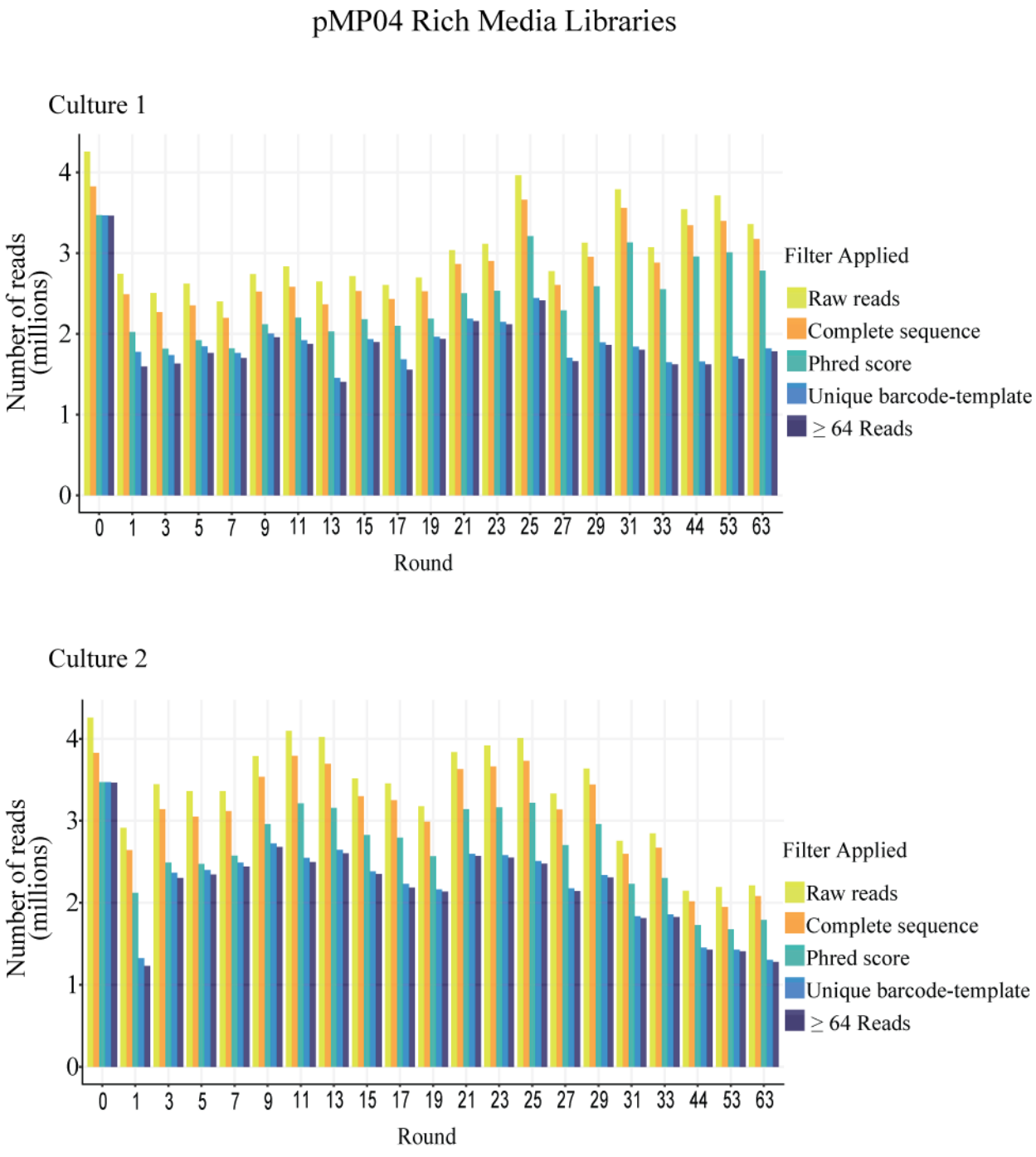


Figure 2.1 Filtering scheme and final read counts for the pMP04 Rich Media libraries. Number of reads after each filter step. Each Round is a single library. Filter steps for each library are organized on the x-axis in the order in which they were performed. The minimum read filter of at least 64 reads was the final filter step and marks the number of reads used for analysis.

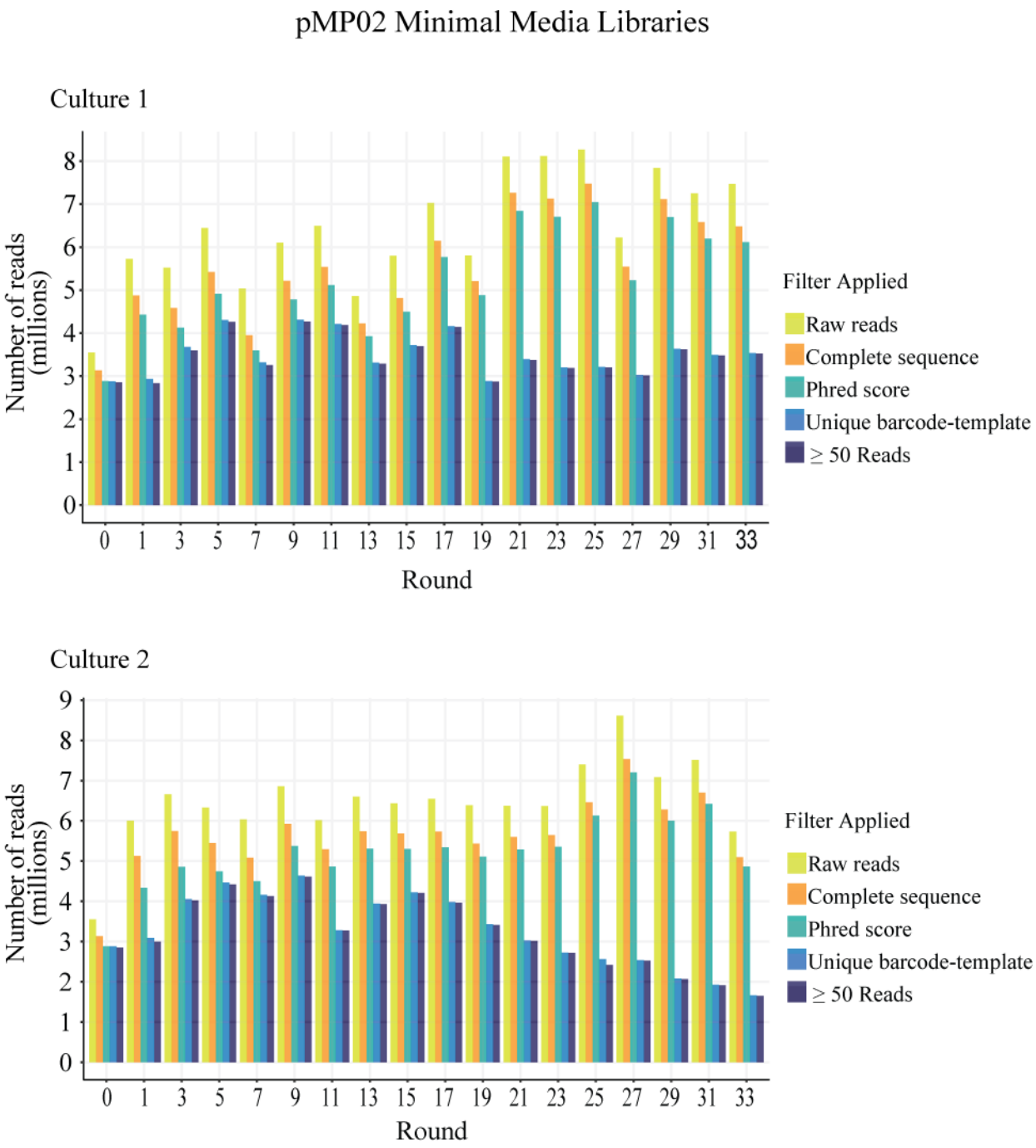


Figure 2.2 Filtering scheme and final read counts for the pMP02 Minimal Media libraries. Number of reads after each filter step. Each Round is a single library. Filter steps for each library are organized on the x-axis in the order in which they were performed. The minimum read filter of at least 50 reads was the final filter step and marks the number of reads used for analysis.

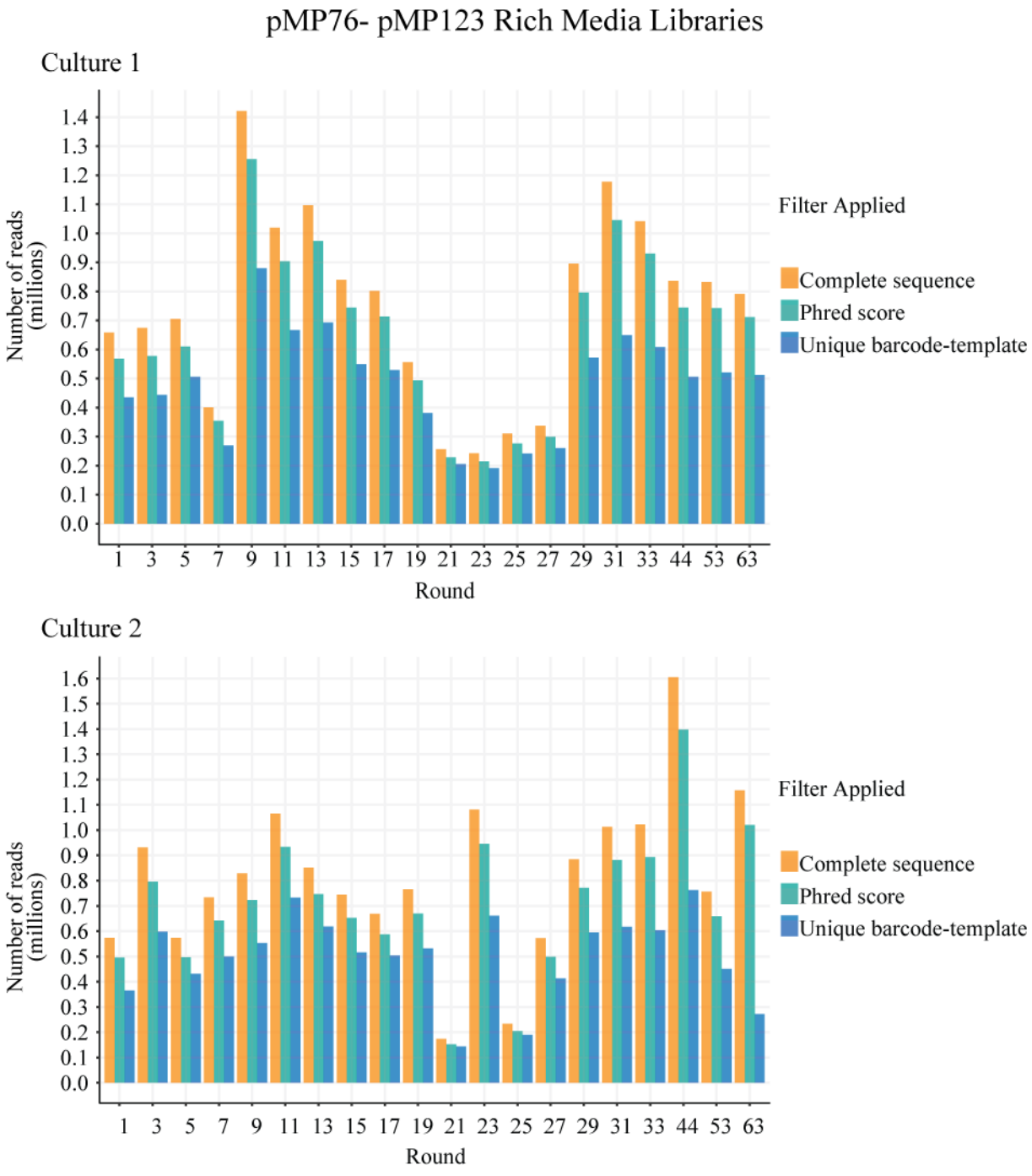


Figure 2.3 Filtering scheme and final read counts for the pMP76 – pMP123 Rich Media libraries. Number of reads after each filter step. Each Round is a single library. Filter steps for each library are organized on the x-axis in the order in which they were performed. Raw read counts were omitted because overlapping indices were used to pool these custom libraries with more complex genomic DNA libraries. No minimum read filter was used. The retention of unique barcode-template pairs was the final filter step and marks the number of reads used for analysis.

Calculations were performed as follows. All read counts were normalized to the total number of reads at a specific time point. Population doublings (H′) were determined using Equation 1 where R represents the Round, n denotes the number of the final round, and C denotes the concentration at time point n.

Equation 1: $$H' = \sum_{R1}^{Rn} \frac{ln\left(\frac{C2}{5x10^5}\right)}{ln(2)} + \cdots + \frac{ln\left(\frac{Cn}{5x10^5}\right)}{ln(2)}$$

The frequency of the wild type template from the sequencing data was used to determine the number of wild type generations (G′, Equation 2) [206].

Equation 2 $$G' = \sum_{R1}^{Rn} \log_2 \left(\frac{normalized\ wild\ type\ reads\ Rn}{normalized\ wild\ type\ reads\ R1} x \frac{C2}{5x10^5}\ x \ldots x \frac{Cn}{5x10^5}\right)$$

The normalized change in frequency is the ratio of frequencies of a given template between Round 1 and a specified time point (n) compared to the ratio of frequencies for wild type (Equation 3).

Equation 3 $$Normalized\ frequency\ in\ change$$

$$= (\frac{normalized\ reads\ of\ template\ Rn/normalized\ reads\ of\ template\ R1}{normalized\ wild\ type\ reads\ Rn\ /normalized\ wild\ type\ reads\ R1}$$

The relative fitness (F) of a template was calculated by combining the number of wild type generations with the normalized change in frequency (Equation 4) [206].

Equation 4 $$F = \left(\frac{normalized\ template\ reads\ Rn/normalized\ template\ reads\ R1}{normalized\ wild\ type\ reads\ Rn/normalized\ wild\ type\ reads\ R1}\right)^{1/G'}$$

II.3.6 Statistical Analysis

To determine the significance of changes in cytosine content between templates prior to and after selection, the Multinomial Goodness of Fit test was used. To determine the distribution of normalized change in frequency for 3'-NCCAAUN-5', 3'-NNCCAAU-5', and 3'-

CCAAUNN-5' templates, graphs were constructed using the boxplot function in ggplot2 Software. The upper bound of the box represents the third quartile (75th percentile) and the to-most point of the upper whisker corresponds to the highest value within 1.5 times the distance between the first and third quartiles [207]. The lower bound of the box represents the first quartile (25th percentile) and the lowest point of the lower whisker corresponds to the lowest value within 1.5 time the distance between the first and third quartiles [207]. Outliers are defined as the points outside of the end of the whisker [207].

II.4 Results

II.4.1 Population dynamics in a competitive growth time course

To probe the range of telomeric sequences that can maintain cell growth, a plasmid library (pMP04) expressing TER1 sequences with seven randomized nucleotides within the template was generated (Figure 2.4A). Haploid cells lacking the endogenous *ter1* gene were transformed with the plasmid library and grown competitively in two independent cultures for sixty-three rounds of dilution and growth to stationary phase (Figure 2.4B). The initial library (R0) demonstrated a nearly normal distribution with four overrepresented templates 3'-AAAUUGG-5', 3'-CUCAAAA-5', 3'-UCCUUAA-5', and 3'-UCUCAUG-5' (Figure 2.4C). However, the abundance of the wild type template was within the normal distribution. While the initial library contained all 16,384 possible nucleotide combinations of a randomized seven nucleotide sequence, only 76-172 templates were identified between the two cultures by the final time point (Figure 2.4D). Additionally, five to eight templates accounted for 90% of reads by Round 63. The total number of generations of wild type strains in culture was determined from the frequency of the wild type template and culture densities after each round. The total number of population doublings was determined directly from culture density counts after each round. Surprisingly, the total number of population generations closely approximated the total number

of wild type generations, indicating that most cells in the population were growing as well as wild type (Figure 2.5A). More specifically, by Round 33, wild type strains had undergone 227.9 and 236.5 generations in Culture 1 and 2, respectively, compared to 230.1 and 233.0 population doublings, respectively. By the final time point, Round 63, wild type strains had completed 454.3 and 463.1 generations in Culture 1 and 2, respectively, compared to 455.2 and 460.6 population doublings, respectively. When culture densities were considered, cultures started out with slow growth, barely reaching stationary phase within a single round. By the later time points, cultures had gradually begun to reach late stationary phase within a single round (Figure 2.5B). This trend is consistent with a short telomere phenotype of the overall population in early rounds that gradually increased over time (Figure 2.5C). Interestingly, by the final round of 63, after 455-460 population doublings, the pattern of cytosine (C's) content among the remaining templates was significantly different ($p < 2.2x10^{-16}$) from the library prior to competitive growth and selection. Templates favored a pattern of two consecutive C's (Figure 2.5D).

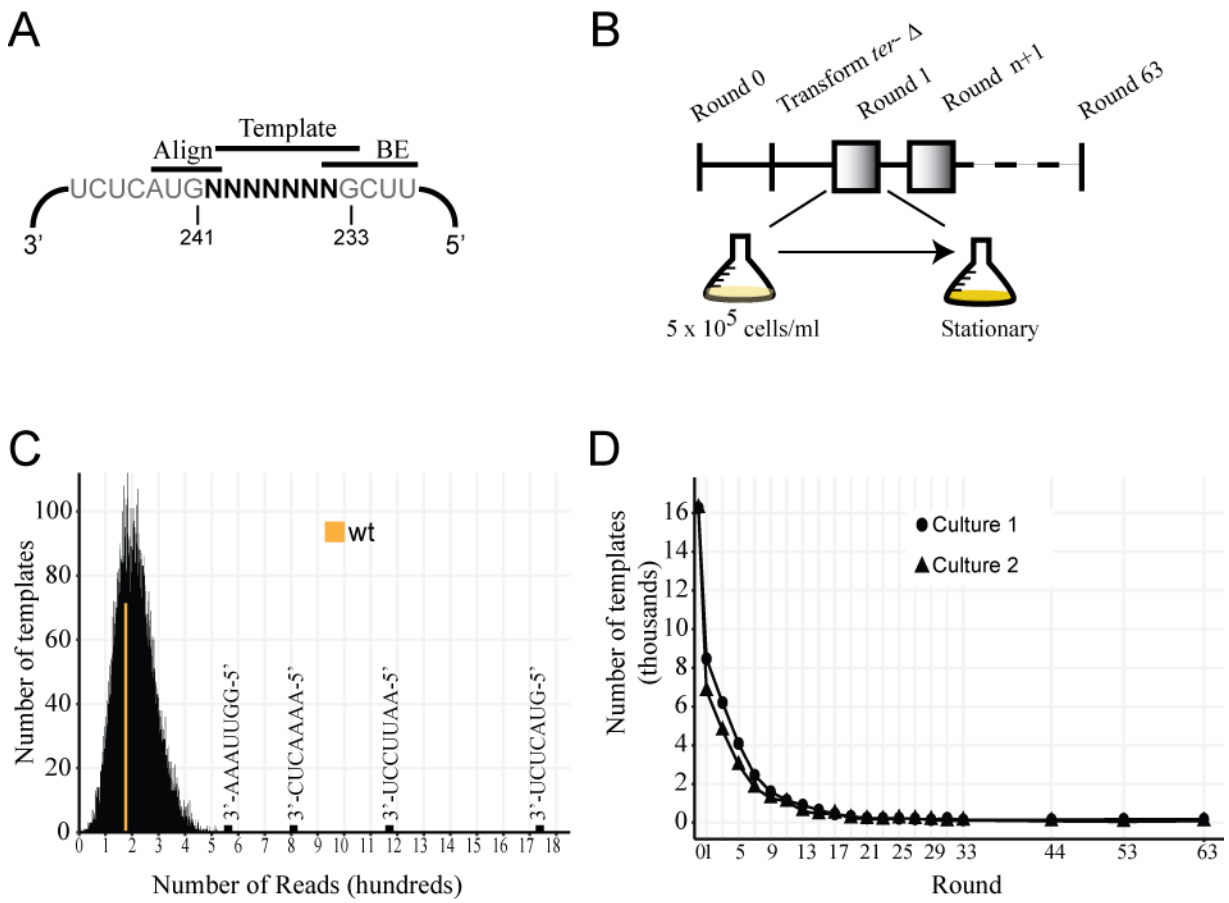


Figure 2.4 Decrease in TER1 template library diversity throughout competition time course. (A) Randomized nucleotides denoted with a bolded N. The alignment, template, and boundary elements (BE) of the wild type TER1 RNA are shown. (B) Design of the in flask evolution experiment. Plasmid library (pMP04) was introduced into presenescent haploid *ter1*$^{-}$ cells sporulated from PP407 and used for transformation. Cells were grown in Rich Media with 50μg/ml ClonNat. (C) Distribution of template sequences in the Round 0 library prior to transformation into yeast. Wild type (wt) template is highlighted in orange. (D) Number of templates per round is shown for two independent cultures.

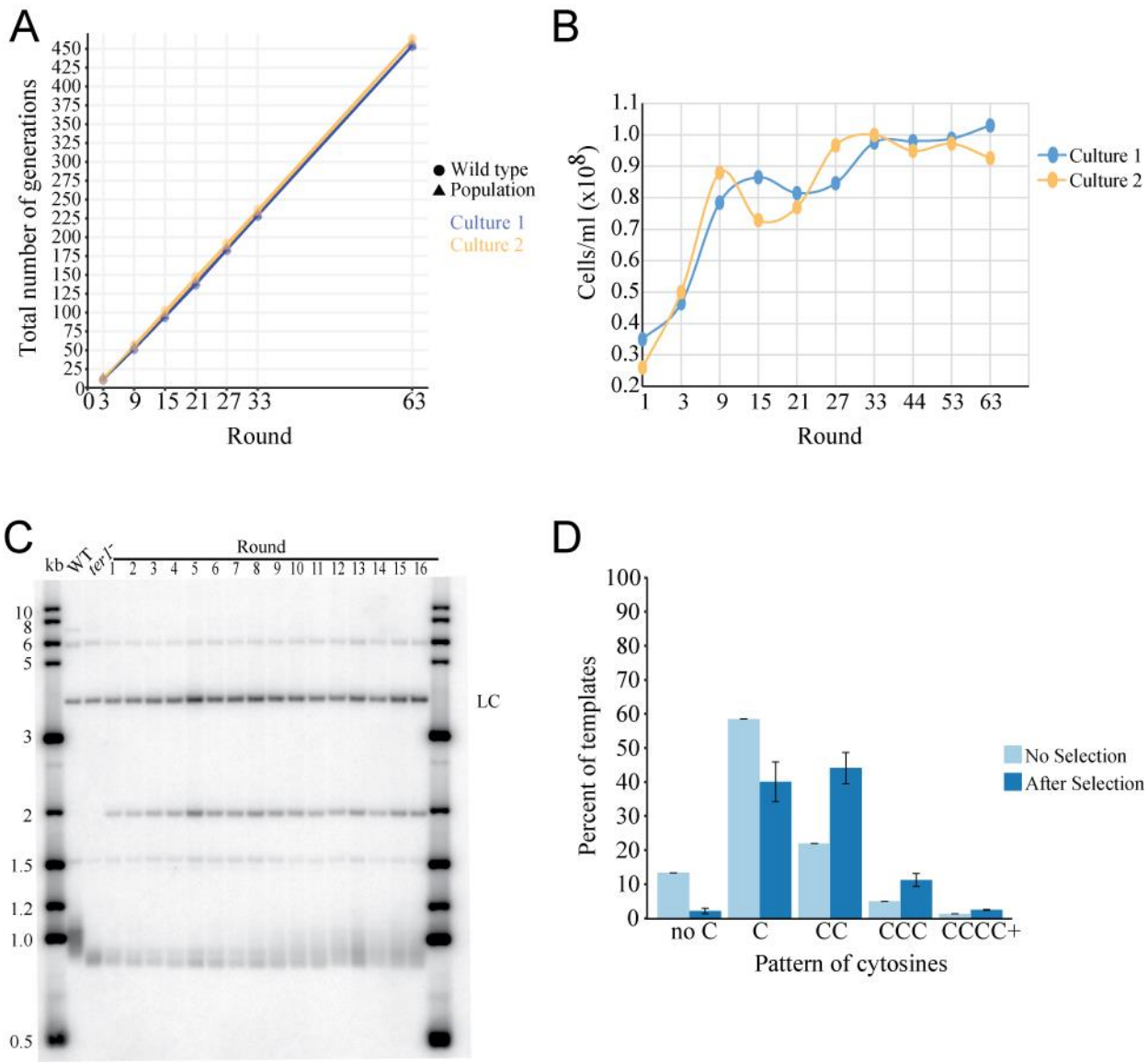


Figure 2.5 Population telomere length reflects initial slow growth of cultures. (A) Cumulative number of wild type doublings and population doublings after each round. (B) Culture densities were determined by counting in a hemocytometer. Early stationary phase ~ $2.6x10^7$ cells/ml; late stationary phase ~$1.03x10^8$ cells/ml. (C) Southern blot with a wild type telomere specific probe; PP138 was used as the wild type (WT) control; *ter1*$^-$ is the PP407 haploid strain prior to plasmid library transformation. A rad16-specific probe was used as the loading control (LC). (D) CCCC+ represents sequences of 4 or more consecutive cytosines. After Selection: percentages are calculated as the fraction of templates with the specified pattern of consecutive cytosines out of the total number of unique templates at Round 63 multiplied by 100. No selection refers to the theoretical percentage of templates if all 16,384 templates were present. Error bars represent the mean +/- standard error.

II.4.2 A 3'-CCAAU-5' is the common feature of the most competitive templates

The decrease in the number of templates accounting for 90% of the library and the emerging pattern of cytosine content by Round 63, together indicated a selective process. It was hypothesized that selection could be used to define the core features of a template that supported growth rates comparable to a wild type template strain. Indeed, analysis of nucleotide patterns ranging from three to five nucleotides in length revealed shared patterns among the most abundant templates. To quantify the success of sequences throughout the time course, fitness scores were calculated for all 1,024 possibilities of a five-nucleotide (5nt) sequence excluding read counts from the wild type template. Initially, at Round 3, the vast majority of sequences displayed an apparent fitness greater than wild type (Figure 2.6A). Then by Round 63, 61.4% had reached a fitness of 0.00, 36.7% a fitness between 0 and 1, and 1.9% with a fitness greater than 1. Of the sequences with a fitness between 0 and 1, 26.2% were between a fitness of 0 and 0.9 (light blue) and 10.5% were between 0.9 and 1 (lighter blue) (Figure 2.6A). The two sequences with the highest average fitness scores were firstly 3'-UCCCA-5' and secondly 3'-CCAAU-5'. A more targeted analysis of 5nt sequences confirmed that the fitness of specific patterns declined to zero at different time points, while the average fitness of the 3'-UCCCA-5' and 3'-CCAAU-5' sequences remained greater than wild type (Figure 2.6B). To distinguish between these two patterns, we considered the frequency of each pattern among templates present at Round 63. The sequence 3'-UCCCA-5' accounted for 25-28% of reads whereas 3'-CCAAU-5' was found in 96-99% of reads. The higher fitness of 3'-UCCCA-5' was the result of lower starting read counts at Round 1 than 3'-CCAAU-5'. Additionally, the most abundant templates at Round 63 all contained a 3'-CCAAU-5' sequence (Figure 2.6C). Clones carrying one of these templates were isolated from Culture 2 at Round 63 and analyzed by Southern blot

for telomere length. As expected, linear telomeres were maintained in all isolates. However, isolates with the templates 3'-**UU**CCAAU-5', 3'-**CG**CCAAU-5', and 3'-**UC**CCAAU-5' displayed slightly shorter telomeres than wild type (Figure 2.6D). Together these findings demonstrated that a core sequence, 3'-CCAAU-5', existed among the most abundant templates.

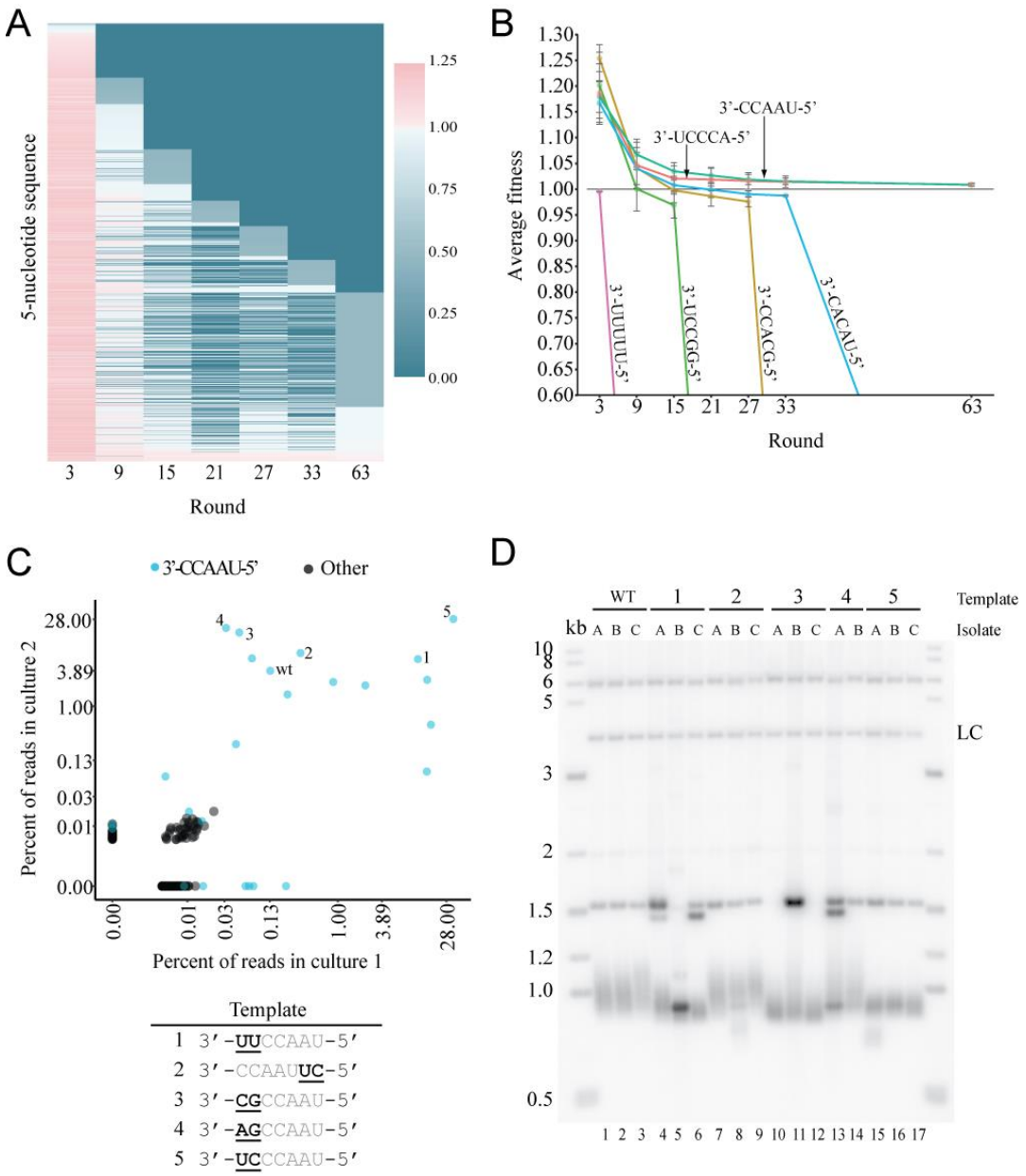


Figure 2.6 A 3'-CCAAU-5' is the most fit 5-nucleotide pattern. (A) Heatmap of the average fitness values between two cultures for each five-nucleotide sequence. Standard error of the mean (SEM) values for the average fitness of each 5nt sequence were below 0.55. A fitness of 1 is equivalent to wild type. (B) The fitness trajectory of specific 5nt patterns. Fitness values were averaged between two cultures. Select templates with SEM values below 0.065 are shown. A black line at 1.00 represents wild type. (C) A comparison of the frequency of entire 7nt templates between two cultures. Each point represents a 7nt template. Axes are shown in a Log (2) scale with percentages rounded to the nearest hundredth. Six templates are labelled as either wild type (wt) or a number corresponding to the template list below the graph. (D) Southern blot of clones isolated from Round 63 of the competition experiment. Blot was hybridized with a subtelomeric element (STE)-specific probe. Templates are labelled with the scheme as in C. A rad16-specific probe was used as the loading control (LC).

II.4.3 Starting position of 3'-CCAAU-5' core within template affects telomere maintenance

Beyond the five nucleotides of 3'-CCAAU-5', the randomized template contained two additional nucleotides. These two nucleotides could be positioned in three different ways with respect to the core 3'-CCAAU-5'. The two nucleotides could be located upstream (3'-CCAAU**NN**-5'), split across the core (3'-**N**CCAAU**N**-5'), or located downstream (3'-**NN**CCAAU-5'). To determine if there was an appreciable difference among these sequence permutations, I repeated the competition experiment with a second library (pMP76 –pMP123) that contained only the 48 seven-nucleotide templates with a 3'-CCAAU-5'. We monitored the success of each template by calculating the ratio or change in frequency from Round 1 to Round 33 or Round 63 normalized to the wild type ratio. Interestingly, 3'-**N**CCAAU**N**-5' templates tended to have lower normalized changes in frequency midway through the time course (Figure 2.7A). However, by the final round, the disparity between 3'-**N**CCAAU**N**-5' and the other two permutations was noticeably diminished, suggesting that cells with these templates started to accumulate in the second half of the time course (Figure 2.7B). The initial slow growth of 3'-**N**CCAAU**N**-5' template strains may be due to an inability to align to the last wild type repeats of the 3' overhang. Plasmid strains expressing a single template were generated to compare telomere phenotypes between cells with the different permutations of the same seven nucleotide sequence. Consistent with an initial growth defect, strains with 3'-**N**CCAAU**N**-5' templates were also found to have critically short telomeres and H1-H1' and H3-H3' end fusions (Figure 2.7C,D). Depending on the two variable nucleotides, 3'-CCAAU**NN**-5' and 3'-**NN**CCAAU-5', template strains were also observed to have short telomeres and end fusions. These telomere defects are consistent with the respective template performance in the competitive growth

experiment, providing further evidence that the position of the 3'-CCAAU-5' core within the greater 7nt template affects telomere length maintenance and cell growth.

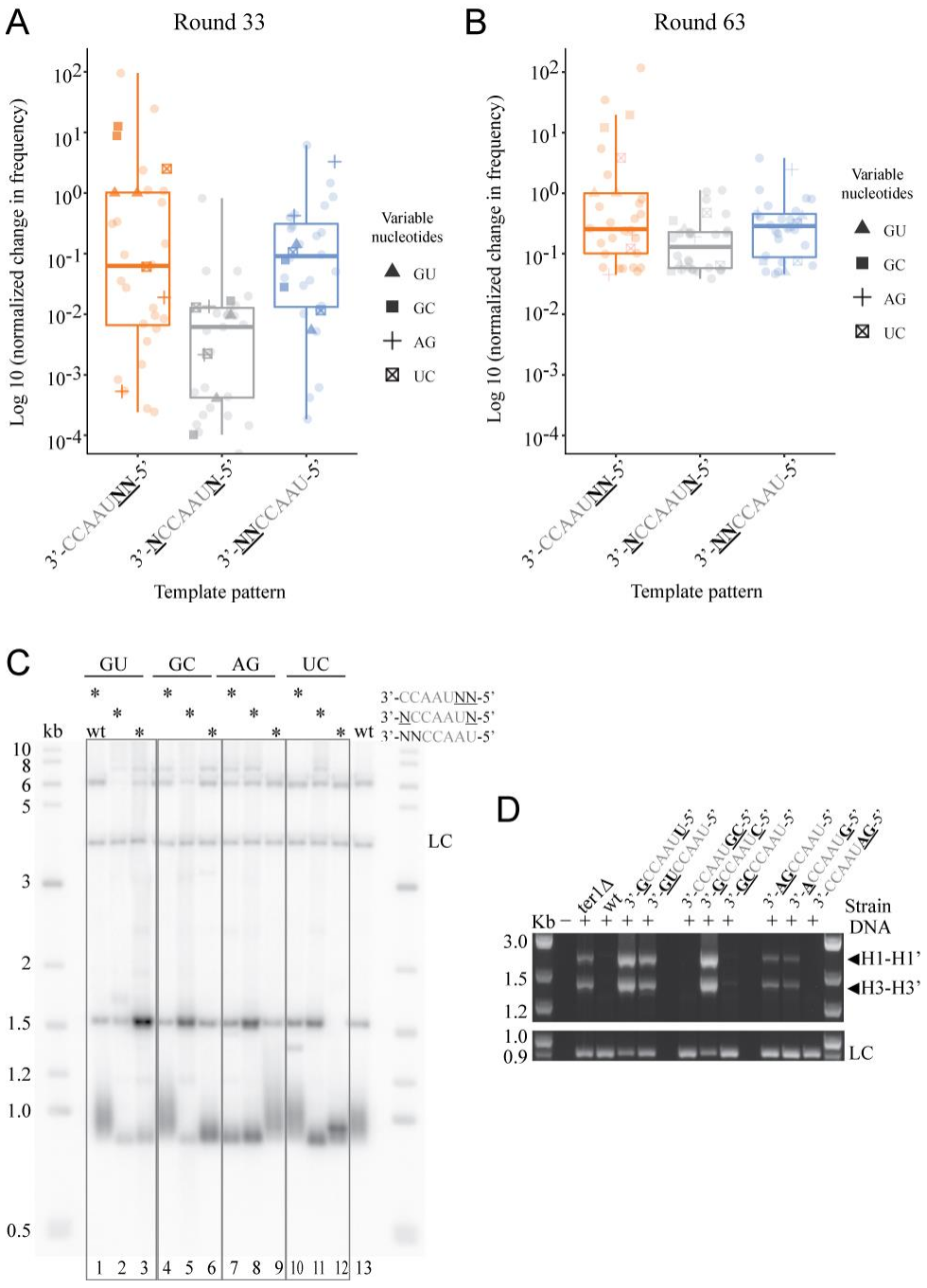


Figure 2.7 The template permutation 3'-NCCAAUN-5' results in short telomeres and end fusions. (A) Normalized change in frequency was calculated as in Equation 3 in the Materials and Methods Section II.3.5. Values from both Culture 1 and Culture 2 are shown. (B) Same calculations as done in (A) with Round 63 values used as the end point. (C) Southern blot with a probe specific to the subtelomeric element (STE); FP1418 carrying wild type *ter1* was used as the positive control (wt). A rad16 specific probe was used as the loading control (LC). (D) PCR to amplify end fusions between homologous regions H1-H1' and H3-H3'. Genomic *trt1* was used as a loading control (LC).

II.4.4 Patterns of competitive templates in Minimal Media

S. pombe growth and chromosome end stability are sensitive to environmental perturbations including temperature, nitrogen source, and the amount of auxotrophic factors. A third competition experiment with a different 16,384 *ter1*$^{+}$ plasmid library (pMP02) for growth in minimal media was used to study the selection of template sequences under different growth conditions. Two cultures were transformed and grown independently for 33 rounds of growth to stationary phase followed by dilution to 5 x 10^5 cells/ml. The two cultures grew almost identically, reaching a total of 227.5 to 227.8 generations by the final time point (Figure 2.8A). Wild type template strains grew at nearly the same rate as the overall population with a total of 226.6 and 226.9 generations in culture 1 and 2, respectively, by the final time point. However, the telomere length of the entire population remained short throughout the time course, signifying the abundance of templates that maintained short linear telomeres (Figure 2.8B). To begin to profile the specific templates in the population, we characterized the cytosine-content of all templates present by round 33 as done for the rich media competition. The cytosine-content had significantly changed ($p < 2.2x10^{-16}$) from the distribution of the original library prior to selection (Figure 2.8C). Templates with two consecutive cytosines were enriched after 33 rounds of selection. Accordingly, even when the wild type template was excluded from analysis, 3'-CCAAU-5' was determined to have an average fitness of 1.05 in cultures 1 and 2, making it the most fit in minimal media among all possible 5nt sequences (Figure 2.8D). Generally, 54.3% of 5nt sequences had fitness values equal to 0 (Dark blue, Figure 2.8D), 43.2% had fitness values between 0 and 1 (Light blue and lighter blue, Figure 2.8D), and 2.5% had fitness values greater than 1 (Light pink, Figure 2.8D).

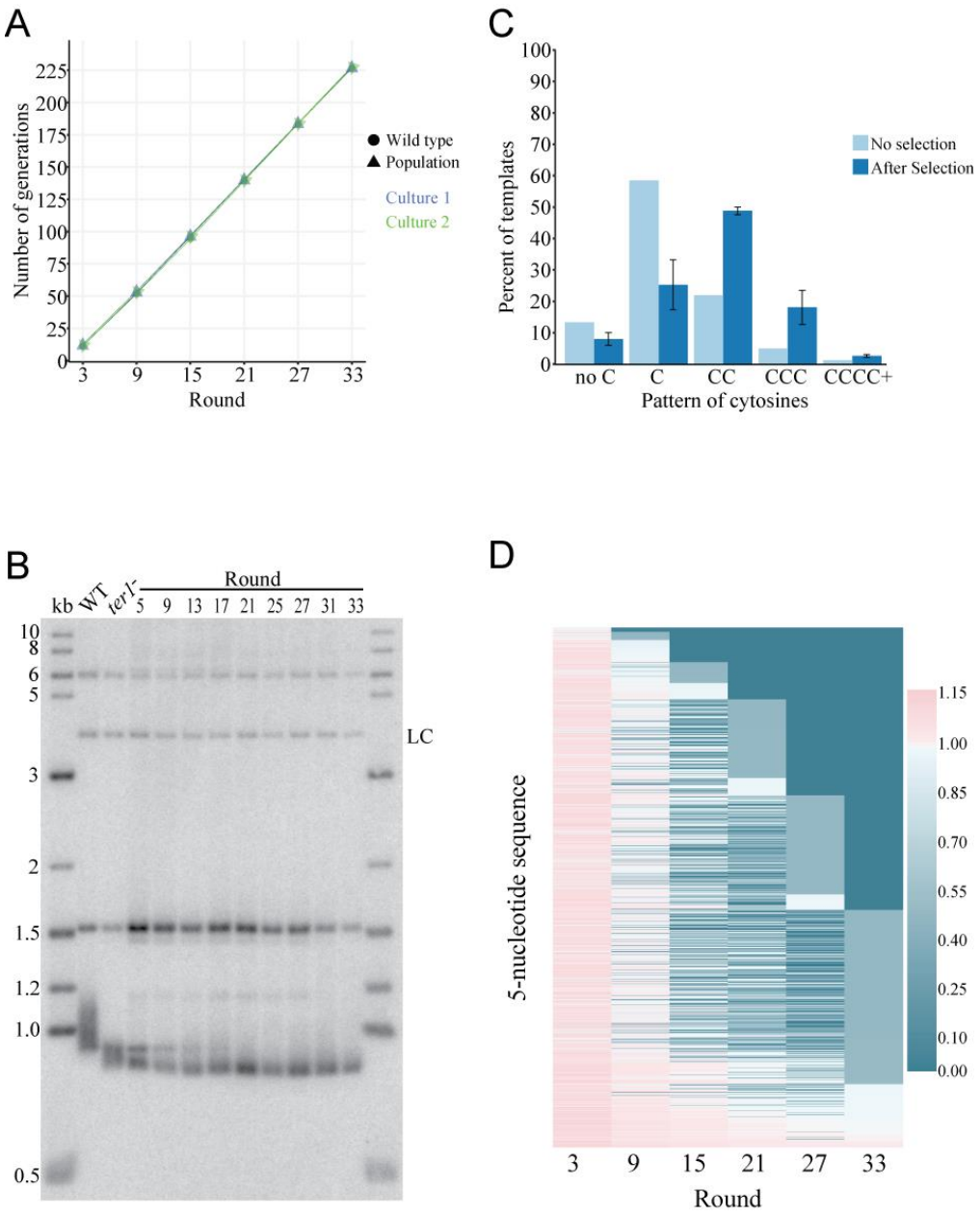


Figure 2.8 Templates with two consecutive cytosines and 3'-CCAAU-5' are the most competitive in minimal media. (A) Cumulative number of wild type doublings and population doublings after each round. Minimal media time course was carried out for 33 rounds. (B) Southern blot hybridized with a probe specific for the subtelomeric element (STE). PP138 was used as the wild type (WT) control. The *ter1*$^{-}$ strain was sporulated from PP407, grown to log phase in rich media, and collected prior to pMP02 plasmid library transformation. (C) CCCC+ represents sequences of 4 or more consecutive cytosines. After Selection percentages are the fraction of templates with the specified pattern of consecutive cytosines out of the total number of unique templates at Round 63 multiplied by 100. No selection refers to the theoretical percentage of templates if all 16,384 templates were present. Error bars represent the mean +/- standard error. (D) Fitness values for each 5nt sequence were averaged between two cultures. Standard error of the mean values were each below 0.52. A fitness of 1 is equivalent to wild type.

II.4.5 Templates lacking the 3'-CCAAU-5' pentamer emerge among competitive strains

While 3'-CCAAU-5' demonstrated the greatest average fitness among 5nt sequences regardless of culture media, the dynamics of selection for this sequence varied between rich and minimal media competitions. By Round 33, the rich and minimal media cultures had undergone a similar number of generations between 227.5 and 233.0 (Figures 2.5A and 2.8A). Yet, despite the similar growth rates, the percentage of 3'-CCAAU-5' containing templates increased at a slower rate in minimal media (Figure 2.9A). By Round 33, 3'-CCAAU-5' was found among 99.1% and 98.6% of reads in Culture 1 and 2, respectively, in rich media. Conversely, only 69.1% and 53.3% of Round 33 Culture 1 and 2 reads, respectively, contained the 3'-CCAAU-5' pentamer in minimal media. The persistence of non-3'-CCAAU-5' templates was further demonstrated by a comparison of 7nt-template frequencies between the two cultures at Round 33 (Figure 2.9B). Two templates lacking a 3'-CCAAU-5' core, 3'-UCCUCGG-5' and 3'-UCCAAGU-5', were among the most frequent templates in culture 1. This finding was consistent with the fitness analysis, which revealed fitness scores of 1.02 in culture 1 for three 5nt motifs within these two templates. Motifs, 3'-CUCGG-5'and 3'-UCCUC-5'are found within the template 3'-UCCUCGG-5' and the motif, 3'-CAAGU-5', is found within the 3'-UCCAAGU-5' template. Telomere lengths were critically short in 3'-UCCUCGG-5' clones isolated from culture 1 at Round 33 (Figure 2.9C). Conversely, telomere length was dramatically increased in the 3'-UCCAAGU-5' clone. Linear telomeres in the clones of each template, 3'-UCCUCGG-5', 3'-UCCAAGU-5', and 3'-GUCCAAU-5' were unstable, resulting in chromosome end fusions between H1-H1' and H3-H3' chromosomal homology regions, respectively (Figure 2.9D). A greater collective number and individual frequency of templates like 3'-UCCUCGG-5' and 3'-GUCCAAU-5' that correspond to short telomeres can explain the

overall short telomere phenotype of the whole population (Figure 2.8B). Together these findings indicate that the minimal media environment is more permissive. While 3'-CCAAU-5' still represents the fittest templates, it is not required for competitive growth in minimal media.

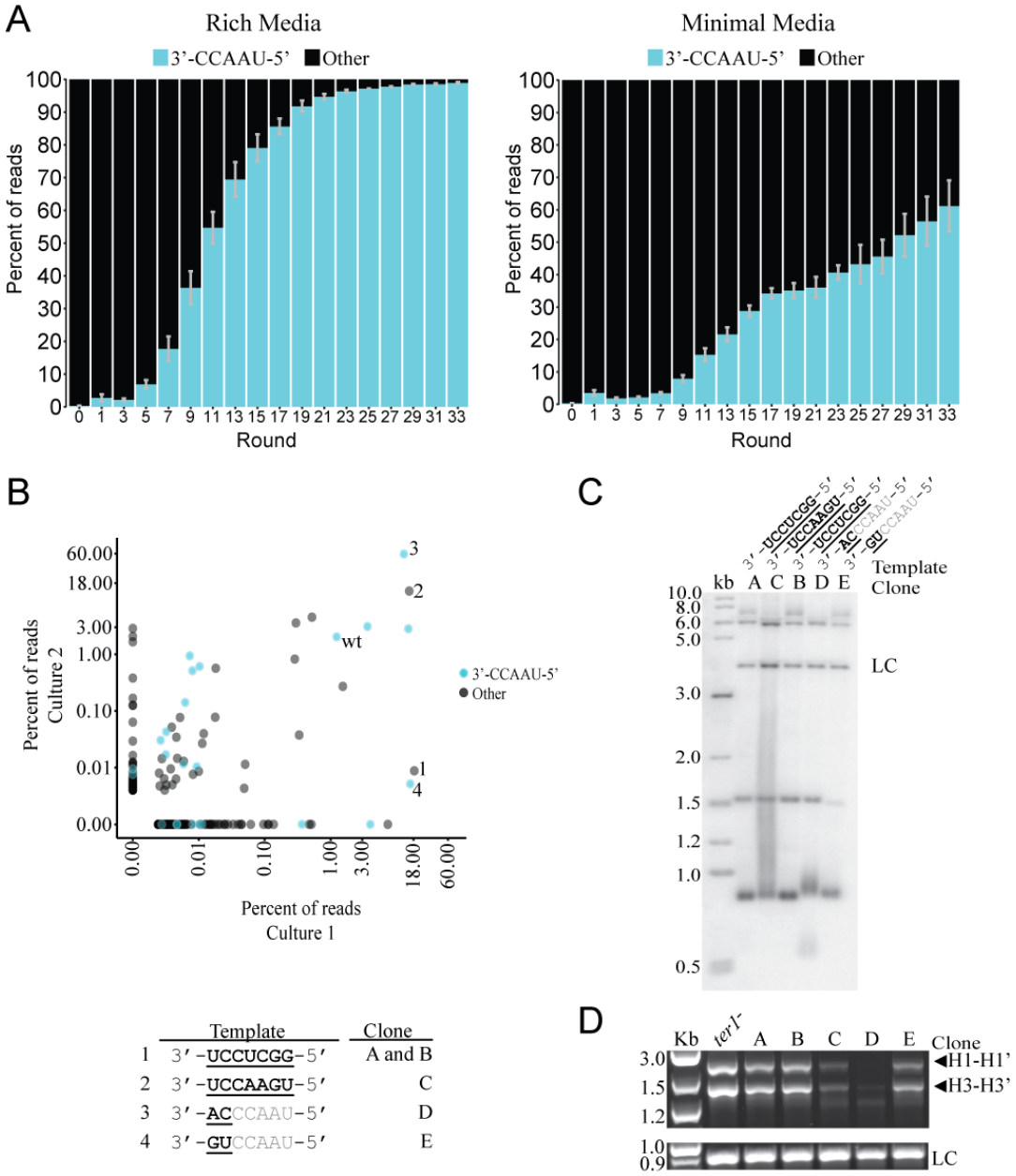


Figure 2.9 The 3'-CCAAU-5' pentamer is not required for competitive growth in minimal media. (A) Comparison between rich and minimal media experiments of the frequency of templates containing 3'-CCAAU-5'. Bars represent the average frequency between two cultures and error bars were calculated as average +/- standard error. (B) Correlation plot comparing the frequency of 7nt templates (single dot) between cultures 1 and 2. Templates are labelled as wild type (wt) or numerically with the corresponding template sequence in the list below the graph. (C) Southern blot of clones isolated from Round 33 culture 1. Clone labelling is consistent with B. A probe specific to the subtelomeric element (STE) was used to detect telomeres and a probe specific to rad16 was used as the loading control (LC). (D) PCR to detect chromosome end fusions between H1-H1' or H3-H3' homology regions. Clone labelling is consistent with B. The catalytic subunit of telomerase *trt1* was used as a loading control (LC).

II.4.6 Telomerase alignment and templated repeat addition explain the growth advantage of template winners lacking 3'-CCAAU-5': Template 3'UCCAAGU-5'

The short telomeres of template strains 3'-UCCAAGU-5' and 3'-UCCUCGG-5' seem to conflict with their competitive growth in liquid culture. To better understand how these templates strains became overrepresented in Culture 1, telomeres from each strain were cloned and sequenced. The 3'-UCCAAGU-5' strain templated 5'-GGTTCA-3' repeats that could be reverse transcribed directly onto the last wild type 5'-GGTTACA-3' repeat (Figure 2.10). The 5'-GGTTCA-3' motif was incorporated on average as three repeats with as many as 14-16 repeats in some telomeres (Figure 2.11A). The C240U mutation in the variant template allows TER1 to form four canonical base pairs with the last four nucleotides of the wild type repeat 5'-GGTTACA-3' (Figure 2.11B). In this way, C_{239} of the variant template becomes the +1 register and 5'-GGTTCA-3' repeats can be added directly following 5'-GGTTACA-3'. A translocation step aligns the last two nucleotides of 5'-GGTTCA-3' to GU_{241} such that the next 5'-GGTTCA-3' repeat can begin in the +1 register (Figure 2.11C).

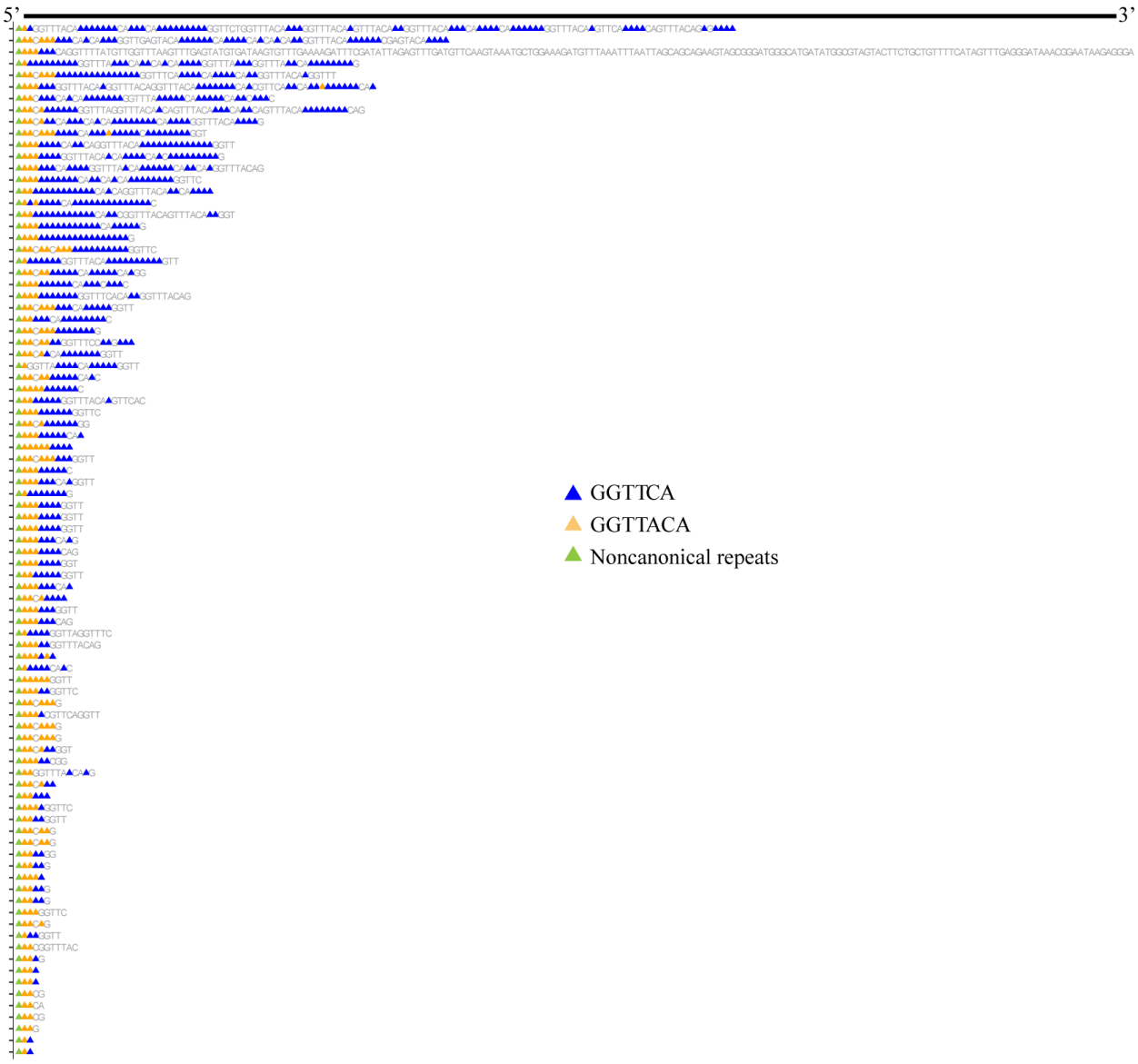


Figure 2.10 Direct addition of variant repeats to the wild type 3' overhang in the template strain, 3'-UCCAAGU-5'. Alignment of all cloned telomeres from the 3'-UCCAAGU-5' clone isolated from Culture 1 at Round 33. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' (gold triangle) are clustered toward the proximal end of the telomere. The most abundant repeat, 5'-GGTTCA-3' is highlighted by a blue triangle.

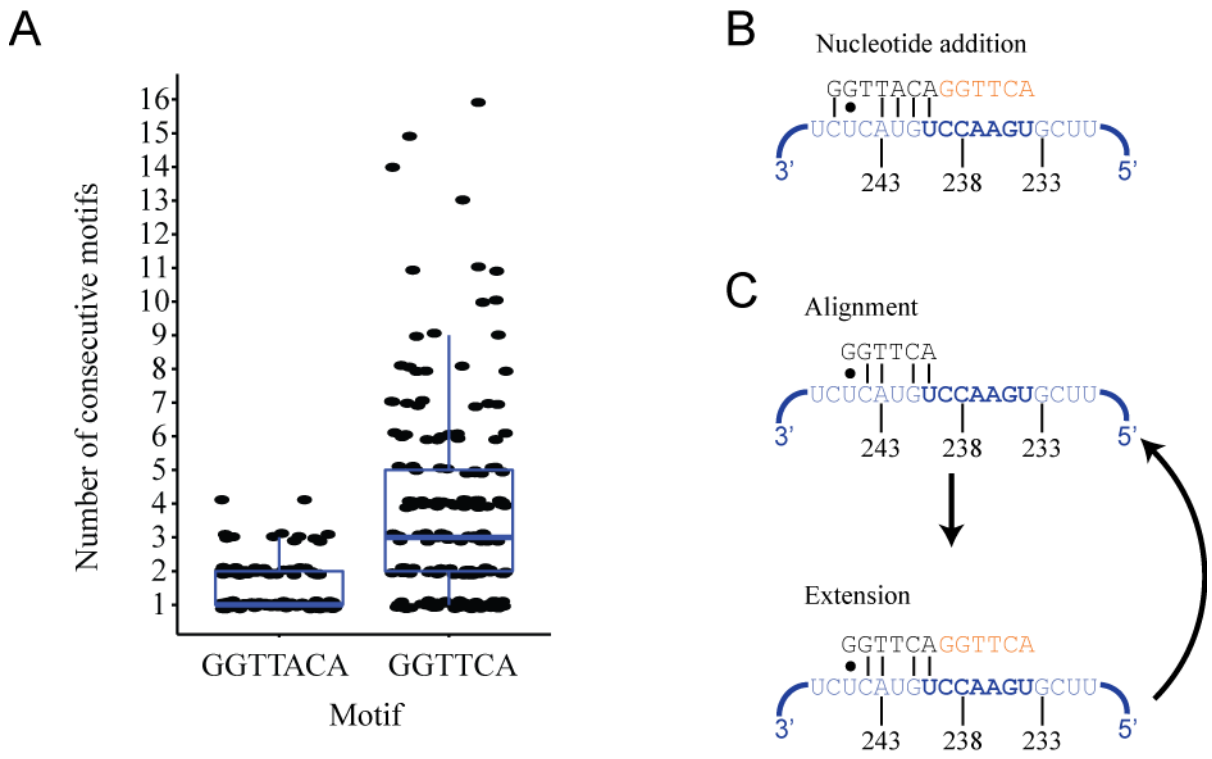


Figure 2.11 Mechanism of direct addition of templated, variant repeats to wild type 3' overhang. Sequences of 89 telomeres (8,802nts combined). (A) Sequencing data were analyzed for motifs that generated the highest number of consecutive repeats throughout the telomere. Stretches of motifs were counted independently. (B) Alignment of variant template TER1 (blue, template is bolded) to wild type 3'-overhang (black). Newly reverse transcribed sequence in orange. (C) Repeat addition of variant 5'-GGTTCA-3' repeats; TER1 (blue, template bolded), 3'overhang (black), canonical base pairs (vertical line), noncanonical base pairs (single dot), newly synthesized repeat (orange). One repeat addition cycle is depicted.

II.4.7 Telomerase alignment and templated repeat addition explain the growth advantage of template winners lacking 3'-CCAAU-5': Template 3'UCCUCGG-5'

Sequencing analysis of the cloned telomeres from template strain 3'-UCCUCGG-5' revealed a 5'-GGAG-3' immediately following the last wild type 5'-GGTTACA-3' in 52 of 79 telomeres (Figure 2.12). Additionally, the 5'-GGAG-3' sequence coincided with the distal four nucleotides of the most abundant variant repeat, 5'-TACAGGAG-3'. The 5'-TACAGGAG-3' motif was incorporated into 2-5 consecutive repeats (Figure 2.13A). Like the 3'-UCCAAGU-5' template, the 3'-UCCUCGG-5'template introduced a C240U mutation into the template region

which facilitated TER1 alignment to the 5'-TACA-3' of the last wild type repeat (Figure 2.13D). This initial alignment resulted in the addition of 5'-GGAG-3' and primed telomerase to use the end of this variant repeat in subsequent alignment steps. Notably, repeat addition of the variant motif is then facilitated by three canonical base pairs between the last three nucleotides of 5'-GGAG-3' and a shifted alignment region, CUC_{244} in TER1. This alignment is shifted 3 nucleotides downstream from the wild type AUG_{241} in TER1, demonstrating a surprising degree of flexibility in the telomerase RNA-telomere interaction.

Motifs lacking an interrupting spacer sequence accounted for 77.0% of repeats (Figure 2.13B). The 8nt sequence, 5'-TCCAGGAG-3', was the most frequent spacer occurring between 14.8% of repeats (Figure 2.13B, C). This spacer includes a 5'-T**C**CAGGAG-3' nucleotide (bolded) that cannot be templated by nucleotides in the variant alignment or template regions. In fact, no matching reverse complement sequence was found in the entire TER1. Telomerases from both human and the ciliate, *P. tetraurelia,* have been shown to included single nucleotide polymorphisms thought to be the result of an error-prone telomerase [208, 209]. Similarly, the 5'-TCCAGGAG-3' spacer sequence may be the result of unfaithful reverse transcription of the TER1 template. Alternatively, a mutation in the telomere arising from replication could also cause the inclusion of this spacer sequence in the telomere.

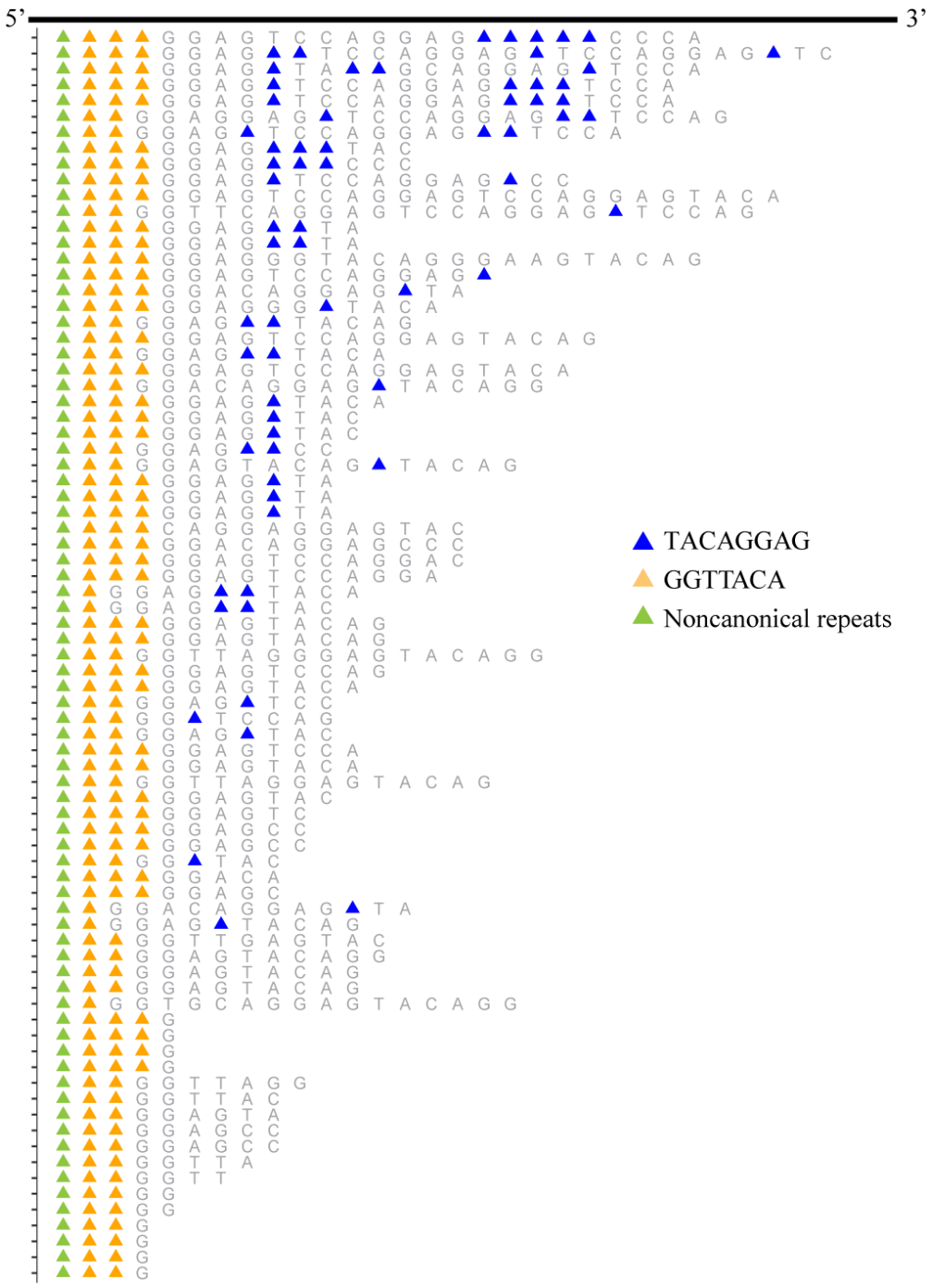


Figure 2.12 5'-GGAG-3' follows the last wild type 5'-GGTTACA-3' repeat and forms the distal half of the variant repeat sequence. Alignment of all cloned telomeres from the 3'-UCCUCGG-5' clone isolated from Culture 1 at Round 33. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' (gold triangle) are clustered toward the proximal end of the telomere. The most abundant repeat, 5'-TACAGGAG-3' is highlighted by a blue triangle.

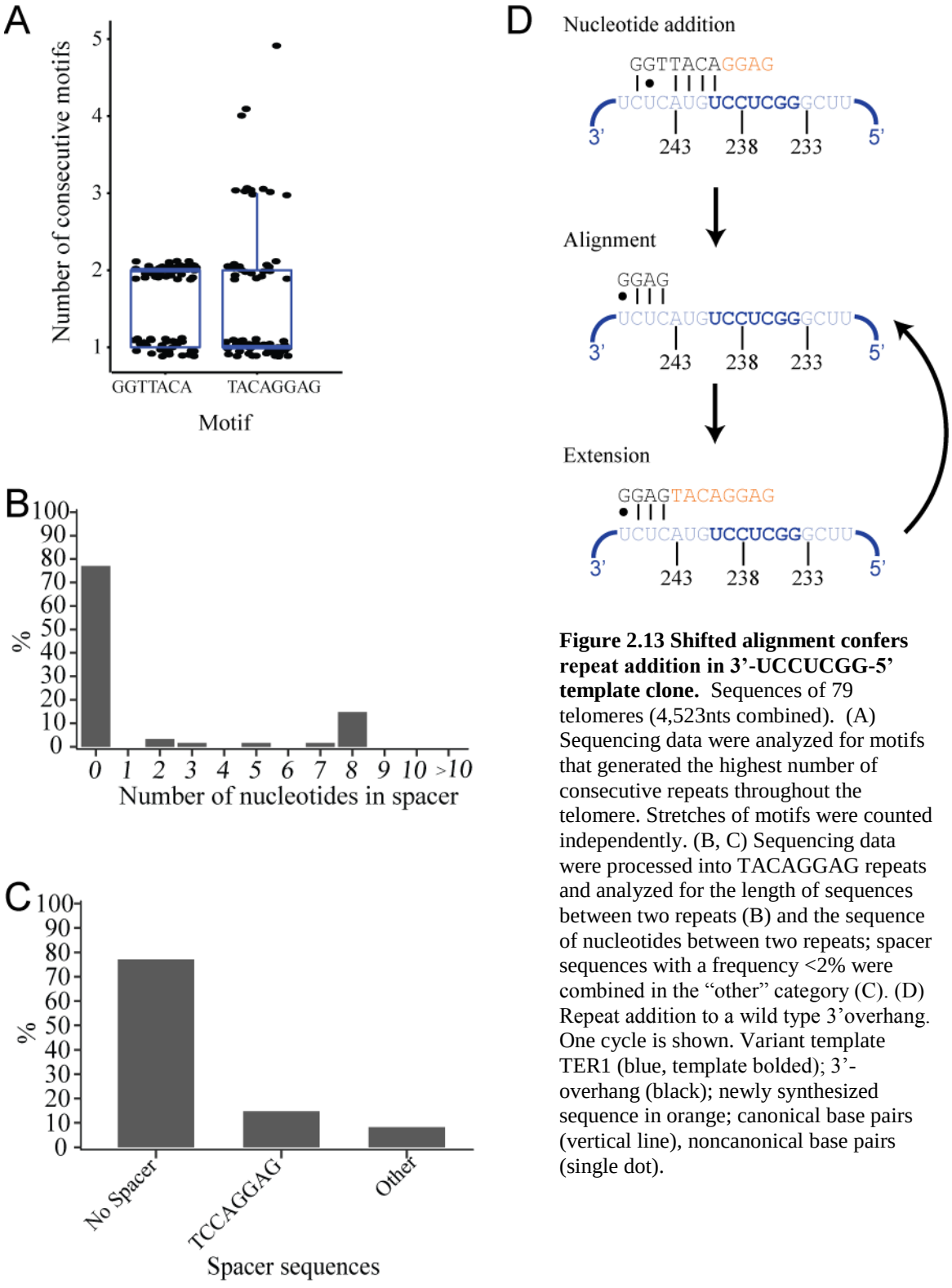


Figure 2.13 Shifted alignment confers repeat addition in 3'-UCCUCGG-5' template clone. Sequences of 79 telomeres (4,523nts combined). (A) Sequencing data were analyzed for motifs that generated the highest number of consecutive repeats throughout the telomere. Stretches of motifs were counted independently. (B, C) Sequencing data were processed into TACAGGAG repeats and analyzed for the length of sequences between two repeats (B) and the sequence of nucleotides between two repeats; spacer sequences with a frequency <2% were combined in the "other" category (C). (D) Repeat addition to a wild type 3'overhang. One cycle is shown. Variant template TER1 (blue, template bolded); 3'-overhang (black); newly synthesized sequence in orange; canonical base pairs (vertical line), noncanonical base pairs (single dot).

II.5 Discussion

II.5.1 Summary

Three iterations of the competitive growth experiment were performed to glean the diversity of template and telomere sequences that can be maintained in *S. pombe*. There was a significant difference between the pattern of cytosines within the template before selection and after selection across experiments. Among the seven nucleotide templates that were present by Round 33 in minimal media and Round 63 in rich media, there was a decrease from the theoretical frequency of templates without a cytosine or with a single cytosine, whereas templates with two or more consecutive cytosines were strongly enriched over the theoretical values. These results are consistent with the natural phenomena of wide-spread G-rich telomeres and provide further evidence that the G-rich telomere confers a competitive advantage.

Additionally, 3'-CCAAU-5' was determined to be the most competitive pentanucleotide pattern in both rich and minimal media. This sequence corresponds to nucleotides known to be required and sufficient for Taz1 and Pot1 binding. Variation of the two nucleotides beyond this core in the template was observed among competitively growing clones in the competitions that started with all seven nucleotide randomized templates. However, when only 3'-CCAAU-5' containing template strains were grown competitively, templates with the 3'-NCCAAUN-5' permutation were associated with an initial growth defect compared to 3'-NNCCAAU-5' and 3'-CCAAUNN-5' templates. 3'-NCCAAUN-5' templates often resulted in impaired base pairing between the mutant telomeric 3' overhang and the template. Together these findings suggest that protein binding is the driving selective pressure in this experimental system followed by a second level of selection from telomerase alignment.

Lastly, we report the discovery of two template clones, 3'-UCCUCGG-5' and 3'-UCCAAGU-5', that grew competitively in minimal media despite lacking nucleotide sequences important for Pot1 and Taz1 binding. Consistent with an inability to bind Pot1, 3'-UCCUCGG-5' template strain telomeres were critically short and unstable. The template strain 3'-UCCAAGU-5' displayed massively elongated telomeres as well as end fusions indicating an inability to bind both Taz1 and Pot1. Surprisingly, 3'-UCCUCGG-5' used an alignment region shifted four nucleotides from the wild type region to facilitate templated nucleotide addition. This is the first time that an alignment shifted this far from the previously defined region has been described, demonstrating remarkable flexibility in the telomere-telomerase RNA interaction.

II.5.2 A prospective approach to telomere evolution

Telomere evolution has primarily been studied through comparative analysis of existing telomere features among diverse organisms. Such studies in budding yeast [59], algae [210], and vertebrates [48] have been used to reconstruct telomere sequence phylogeny and identify conserved features. Comparative biology has also been used to understand specific properties of the telomerase RNA subunit including stability, folding structures, and regions involved in RNA processing and telomerase assembly [37, 38, 211, 212]. Lastly, several telomere binding proteins have been characterized through homology searches for conserved domains including Pot1 [71], TRF1[213, 214], TRF2 [66, 67], and Rap1 [215]. These studies are retrospective in that they focus on events that have already occurred in evolutionary history resulting in the wild type traits that we observe today. Here we employ a prospective evolution experiment that starts with all possibilities of a randomized telomerase RNA template in *S. pombe* and monitors the change in template frequency over time. This prospective approach enabled the selection of telomerase

RNA templates under different growth conditions. The study of these selected templates has demonstrated the impact of variant templates on telomere function and cell growth, providing evidence for how template mutations affect fitness. Lastly, this study has provided insight into how telomerase can adapt to reverse transcribe variant telomerase RNA templates.

II.5.3 Wild type is not the ultimate winner

Prospective evolution studies in budding yeast have also evolved strains with a fitness greater than wild type [206, 216]. In the current study, the wild type telomerase RNA template was not the most frequent in any of the three competition experiments, rendering the first examples of competitive growth among variant telomere strains in *S. pombe*. To understand this observation, several factors should be considered. *S. pombe* exhibits starvation-induced memory which can affect spindle pole body dynamics and microtubules [217]. Since cultures were grown to stationary phase, wild type cells may recover from stationary phase at a slower rate than variant template clones. The experimental conditions may have favored alternative templates in ways that the natural environment would not have. For example, wild type may prove to be the most advantageous when carbon source, nitrogen source, pH, or temperature are varied or if different stressors like hydroxyurea-induced replication stress, ionizing radiation-induced double-strand breaks, or hydrogen peroxide-induced oxidative stress are applied. On the other hand, alternative templates may not be observed in the wild because of an inability to complete meiosis and sporulation. Tetrad dissection of sporulated alternative template strains would provide information on the sporulation efficiency of these strains. Lastly, natural evolution and our in flask evolution are characterized by peaks and valleys in the mutational landscape of a sequence. In this way, any given time point is a snapshot of the bigger picture. Therefore, alternative templates may have arisen at some point in nature prior to the fixation of

the wild type sequence. Interestingly, the wild type template was often the most frequent sequence in early time points (< Round 5). Of note, templates that outcompeted wild type by the final time points had fitness values between 1.01 and 1.04, indicating a very subtle difference between wild type and the alternative template.

II.5.4 Conservation of the 3'-CCAAU-5' telomerase RNA template core

Whereas previous studies have characterized conserved features of wild type telomerase RNA template sequences among eukaryotes [41], we specifically analyzed conserved features among TER1 templates in competitively growing *S. pombe* strains. We found that 3'-CCAAU-5' was the defining feature shared among the most competitive templates in rich media. Consistent with the idea that 3'-CCAAU-5' is important for telomere function and competitive growth, this sequence accounts for five nucleotides of the wild type template that are reverse transcribed into every repeat [60, 61]. Furthermore, this specific sequence is found in templates of vertebrates, echinoderms, filamentous fungi, and other fission yeast [41, 218]. In *S. pombe*, the 3'-CCAAU-5' template sequence is reverse transcribed into telomeric 5'-GGTTA-3', which contains nucleotides known to be important for Pot1 and Taz1 binding (Chapter I.3). Thus, the 3'-CCAAU-5' core is implicated in sequence-specific, protective protein binding and may explain why it was selected for among the most abundant template strains.

Beyond this five nucleotide core, the two variable nucleotides represent a remarkable flexibility in the entire seven nucleotide sequence that can be used to template telomeric repeats and maintain cell division. A similar prospective study with the *S. cerevisiae* telomerase RNA also demonstrated a region of the template that could be mutated without causing senescence [219]. However, we additionally demonstrated that the location of the variable nucleotides in

relationship to the five nucleotide core affects telomere length maintenance and chromosome end stability.

II.5.5 Growth conditions and molecular selective pressures

Comparative analysis of existing telomerase and telomere-associated proteins among various organisms is limited in its ability to predict the effect of molecular changes on evolution. Still, researchers have speculated on the order in which telomere evolution has occurred. Some suggest that the establishment of a uniform telomere sequence allowed for the selection of sequence-specific, telomere-associated proteins, placing telomerase evolution before that of capping proteins [3]. Our findings suggest a more complex scenario, where the relative influence of telomerase activity or protective protein binding on the selection of telomerase RNA templates may be affected by growth condition. In rich media, protective protein binding appears to be the driving force for template selection because all of the most abundant templates contain a 3'-CCAAU-5' core that corresponds to telomeric sequences important for Taz1 and Pot1 binding. Among templates that contain this core, those of the form 3'-NCCAAUN-5' tended to have an initial growth defect and critically short, linear telomeres. The weaker success of these templates compared to other 3'-CCAAU-5' templates suggests that impaired telomerase-mediated telomere elongation restricts template sequence secondary to protein binding.

In contrast to rich media where 3'-CCAAU-5' accounted for an average of 98.9% of reads by Round 33, 3'-CCAAU-5' accounted for an average of 61.2% of reads by Round 33 in minimal media. Indeed, two of the most abundant strains in minimal media contained templates that did not have the 3'-CCAAU-5' sequence. The success of these two templates can be explained by their ability to promote telomerase alignment to the last wild type repeat and

subsequent telomere elongation rather than protective protein binding. Analysis of telomeric sequences from clones of these two template strains demonstrated alternative repeats directly added to the last wild type repeat while lacking apparent Taz1 and Pot1 DNA binding motifs.

Consistent with the presence of deprotected telomeres, the chromosome ends of these strains were highly unstable. The 3'-UCCAAGU-5' template strain telomeres were massively elongated, a phenotype reminiscent of *taz1* deletion strains and the 3'-UCCUCGG-5' template strain telomeres were critically short, more consistent with a *pot1* deletion phenotype. In minimal media stationary cultures, glucose and nitrogen become growth-limiting factors [217]. *S. pombe* arrests in G1 of the cell cycle in response to nitrogen starvation [217, 220, 221]. Additionally, in nitrogen-starved cells, G1 arrest of *taz1*$^{-}$ strains results in classical non-homologous end joining (c-NHEJ)-mediated end-to-end fusions [87]. The 3'- UCCAAGU-5' template strain may have arrested in G1 due to limited nitrogen culture conditions. During G1 arrest, if Taz1 was unable to bind the variant repeats, then c-NHEJ would have produced end-to-end fusions. These fusions could explain the massively elongated telomere phenotype observed by restriction fragment analysis. Alternatively, *taz1*- cells exhibit dramatically increased telomere lengths in a telomerase-dependent manner [62]. Therefore, telomeres in 3'-UCCAAGU-5' may have been elongated due to a lack of negative telomerase regulation as a result of uncapping by Taz1. *S. pombe* arrests in G2 of the cell cycle in response to limited glucose levels [217] and Pot1 deletion [99]. Pot1 deletion results in rapid telomere shortening and end fusions mediated by single strand annealing (SSA) between chromosome end homology regions [99]. The critically short telomeres and homology region end fusions in the 3'-UCCUCGG-5' template strain are consistent with a loss of Pot1 binding. To confirm the mechanism of end fusions in these strains, further analysis must involve sequencing junctions of

end fusions as well as pulse field gel electrophoresis to detect circularized chromosomes. Still, the success of these two template strains, 3'-UCCAAGU-5' and 3'-UCCUCGG-5', points to a biased equilibrium in minimal media where the positive effects of telomerase-mediated telomere elongation are a stronger determinant than the negative effects of telomere deprotection.

Alternatively, the telomeric sequences of the variant template strains may recruit non-canonical proteins that can functionally replace Taz1 and Pot1 in capping the telomere. Studies in fission and budding yeast as well has human have provided some insight into the reciprocal relationship between changes in telomere sequence and recruitment of various species-specific capping proteins. [61, 215, 222, 223]. Our study presents an exciting opportunity to investigate the co-evolution of telomere sequence and binding proteins in real time. Additional studies are needed to identify proteins at the telomeres of these variant strains.

II.5.6 Selection of G-rich telomeres

The telomeres of most organisms are guanine-rich (G-rich) on the 3' overhang strand [23, 224]. The G-rich sequence is conferred by a cytosine (C-rich) template in the telomerase RNA. Based on a phylogenetic study of conserved and divergent telomeric repeat sequences among eukaryotes, researchers have proposed that the ubiquitous nature of G-rich repeats is the result of a common ancestor that harbored TTAGGG repeats [210]. In the present study, we tested whether a C-rich telomerase RNA template would be selected for when beginning with all 16,384 possibilities of a randomized template. Our results strongly indicate that templates with two or more C's are specifically enriched over templates without a C or with a single C. Our findings are supported by a similar prospective study in budding yeast, which also demonstrated a robust preference for two or more consecutive C's in the RNA template [219]. Throughout natural evolutionary history, deviation from a G-rich telomere may have been restricted by

telomere-capping proteins, regulatory folding structures such as the G-quadruplex, and/or telomerase activity. However, the functional role of each of these factors in determining a specifically G-rich telomere requires further investigation. Importantly, our study provides a foundation for studying the contributions of these factors on template and telomere sequence selection in a model organism with telomere biology most closely conserved with vertebrates.

Chapter III: Loss of telomere repeat heterogeneity in *S. pombe*

III.1 Abstract

Telomeres are the deoxyribonucleoprotein structures at chromosome ends that control replicative potential. In contrast to the telomeres of most other eukaryotes, fungal telomeres are generally highly heterogeneous sequences comprised of degenerative repeats. Whether such repeats provide a distinct growth advantage and how homogeneous sequences might be incorporated into fungal telomeres is unknown. In the current study, we cloned and sequenced telomeres from *S. pombe* strains expressing variant telomerase RNA (TER1) templates with the 3'-CCAAU-5' core previously (Chapter II) determined to be important for competitive growth. Here we identify six template strains that incorporate nearly perfect telomeric repeats and display differences in competitive growth. Each of these strains synthesize repeats ending in 5'-TAC-3' from a 3'-AUG-5'introduced into the template. The 5'-TAC-3' ending strongly base pairs with the 3'-AUG-5' of the TER1 alignment through a process reminiscent of vertebrate telomerases. Remarkably, one template strain achieves templated, nucleotide addition with the use of an alignment region shifted four nucleotides downstream of the wild type region. Our results indicate robust repeat addition catalytic activity *in vivo* based on the incorporation of long stretches of successive repeats in cloned telomeres from alternative template strains. Together, these findings challenge the current view that *S. pombe* telomerase displays only nucleotide addition processivity.

III.2 Introduction

The guanine-rich and repetitive features of telomeres are conserved across eukaryotes. Yet, fungi represent a distinct class of telomeres that consist of highly heterogeneous sequences with core repeats interrupted by spacer nucleotides (Chapter I.2). In fission yeast, the wild type repeat is defined as $G_{0-4}GGTTACA_{0-1}C_{0-1,}$ where GGTTAC represents the core found in all repeats. In contrast, vertebrate repeats are largely homogeneous and made up almost entirely of tandem GGTTAG sequences. The differences in telomere sequence are a result of variations in the telomerase RNA template sequence of fission yeast and vertebrate telomerases. The prevalence of heterogeneous telomeres among fungi has led to the assumption that fungal telomerases are inherently non-processive. However, why *S. pombe* telomerase would adopt a more complicated, multistep process for nucleotide addition and repeat synthesis remains unclear. One possibility, that has yet to be experimentally tested, is that homogenous telomeres may be disadvantageous in fungi. In Chapter II, we reported that the five nucleotide sequence, 3'-CCAAU-5', provided the greatest growth advantage over all other five nucleotide sequences. In this chapter, we analyze the telomere sequences of strains harboring templates with the 3'-CCAAU-5' core and two variable nucleotides in order to address these issues in telomere sequence evolution.

III.3 Materials and Methods

III.3.1 Constructs and Strains

Alternative template plasmid constructs were made by cloning as described previously in Chapter II. Briefly, a partial oligonucleotide duplex corresponding to positions 214 to 302 of the *ter1* gene was ligated into pMP75 (Chapter II) and cloned using XL10-Gold Ultracompetent Cells (Agilent Technologies). Bacterial colonies were screened for all 48 plasmids containing 5'-TAACC-3' in the seven nucleotide template including pMP76, pMP77, pMP78, pMP79,

pMP82, pMP85, pMP86, pMP90, pMP94, pMP101, pMP106, pMP110, pMP117, and pMP122. PP433 *ter1*$^{-}$ spores were transformed with each plasmid individually by electroporation. Yeast strains were verified by sequencing and selection on YEA + 100μg/ml Geneticin. The genotypes of the resulting yeast strains are listed in Table 3.1.

Table 3.1 Strains used in this study

Name	Genotype	Source
FP1418	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP78]*	This study
FP1421	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP117]*	This study
FP1430	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP82]*	This study
FP1431	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP90]*	This study
FP1432	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP94]*	This study
FP1433	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP101]*	This study
FP1435	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP106]*	This study
FP1436	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP76]*	This study
FP1437	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP77]*	This study
FP1438	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP79]*	This study

FP1461	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP85]*	This study
FP1462	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP86]*	This study
FP1463	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP110]*	This study
FP1464	*h? leu1-32 ura4-D18 his3-D1 ade6-M216 ter1::ura4 [pMP122]*	This study
PP407	*h+/h- leu1-32/leu1-32 ura4-D18/ura4-D18 his3-D1/his3-D1 ade6-M210/ade6-M216 ter1^{+}/ter1::kanMX6*	Ref. [46]
PP433	*h+/h- leu1-32/leu1-32 ura4-D18/ura4-D18 his3-D1/his3-D1 ade6-M210/ade6-M216 ter1^{+}/ter1::ura4*	Ref. [203]

III.3.2 Genomic DNA Preparation

Cells (~2 x10^9) were collected by centrifugation and washed once with deionized-distilled water (ddH_2O) followed by a wash with 10ml Z buffer (50mM sodium citrate, 50mM sodium phosphate dibasic, 40mM EDTA, pH 7.8). Cells were lysed in 2ml Z buffer, 0.5mg/ml Zymolyase 100T (US Biological), and 2mM dithiothreitol (Sigma) for one hour at 37°C. Then, sodium dodecyl sulfate (SDS) was added to a final concentration of 2.5% (v/v) and the entire mixture was incubated for ten minutes at 65°C. Proteins were digested by adding 5x TE (50mM Tris-HCl, 5mM EDTA, pH 8.0) to a final volume of 10 ml and proteinase K (Sigma) (reconstituted in 50mM Tris pH 8.0, 2mM $CaCl_2$, 10% glycerol v/v) to a final concentration of 50μg/ml, then incubating for one hour at 50°C. Samples were then precipitated by adding 3ml 5M potassium acetate (Sigma) and incubating on ice for thirty minutes. The supernatant was collected after centrifugation at 3700 rpm for ten minutes, mixed with one volume of 100% isopropanol, and incubated on ice for twenty minutes. Excess isopropanol was removed from

precipitated DNA by centrifugation at 7,956 rpm for ten minutes. Then, DNA was solubilized in 5x TE plus 50μg/ml RNAse A (Fermentas, Inc.) and incubated for one hour at 37°C. DNA was extracted twice with 25:24:1 phenol/chloroform/isoamyl alcohol (VWR) equilibrated with 5x TE followed by a single wash with 24:1 chloroform/isoamyl alcohol (Ampresco) equilibrated with 5x TE. Lastly, DNA was precipitated with 2.5 volumes of 100% ethanol for ten minutes on ice, washed with 1 ml of 70% ethanol, and solubilized in 80-100μl 1x TE buffer depending on the size of the pellet.

III.3.3 Spotting Assay

Cultures were grown to log phase ($5 \times 10^6 - 1 \times 10^7$ cells/ml) in YES plus 20μg/ml Geneticin then centrifuged and resuspended at a concentration of 1×10^7 cells/ml. Five-fold dilutions of samples were plated in 5μl spots to YEA 100μg/ml Geneticin in triplicate. Plates were incubated at 32°C and scanned after three days.

III.3.4 Telomere Length Analysis

Genomic DNA was digested for twelve hours at 37°C with EcoRI in 10x EcoRI buffer (New England Biolabs). Restriction fragments were resolved by electrophoresis in a 1% agarose gel in 0.5 x TBE (44.5 mM Tris-borate, 1mM EDTA, pH 8.3) at 160V for 15 minutes followed by 75V for 10-11 hours. After staining the gel for thirty minutes with 1μg/ml ethidium bromide, equal loading was confirmed by imaging the gel with a Typhoon 8600 scanner. DNA was depurinated with 0.25M hydrochloric acid for ten minutes then denatured with 0.5M sodium hydroxide, 1.5M sodium chloride for thirty minutes. DNA was then neutralized by a thirty minute gel wash in 0.5M Tris-HCl, pH 7.5, 1.5M NaCl and transferred by capillary blot to an Amersham Hybond-N+ membrane (GE healthcare Life Sciences) in 10x SSC buffer (1.5M

NaCl, 0.15M sodium citrate, pH 7.0). Each membrane was crosslinked with 120 mJoules of 254 nanometer ultraviolet light.

A probe specific to the subtelomeric element (STE) was amplified by PCR from pTAS1 (lab stock) with the T7 and T3 primers (Table 3.2). The loading control (LC) is a probe specific for the *rad16* gene and generated by PCR of wild type genomic DNA with the XWP9 and XWP10 primers (Table 3.2). Probes were labeled by random-primed labeling with [α-^{32}P]-dCTP and High Prime (Sigma) and purified through Sephadex G25 Fine columns (Sigma). Five million counts per million of each probe were used for hybridization. Membranes were hybridized in Church-Gilbert buffer [225] at 65°C, washed twice for fifteen minutes at 65°C in 0.1% SSC, 0.1% SDS buffer, and then exposed in storage phosphor cassettes (GE healthcare Life Sciences) and visualized with a Typhoon 8600 scanner.

Telomere length was also assayed by counting nucleotides of telomeres cloned using the G overhang Capture Assay. The reported lengths are based on FASTA files where the STE sequence upstream of the noncanonical repeats was trimmed. Therefore, absolute lengths are approximately 792 nucleotides shorter than the southern blot lengths.

Table 3.2 Oligonucleotides used in this study

Name	Sequence
BLoli 1256	5’- GGGTTGCAAAGTATGATTGTGGTAA
BLoli 1353	5’- TGTTGAATGTCAGAACCAACTGTTGCAT
BLoli 3400	5’- GCAAAGAAGTTTCCTGGAATAGC
BLoli 3405	5’- GATGTAATAAAGGGTCGGCAC
PBoli 434	5’- GTGTGGAATTGAGTATGGTGA

PBoli 733	5’- GCGTACGACTCACTGTAGAT NNNNN-3’-O$(CH_2)_2CH_2OH$
PBoli 745	5’- GCGTACGACTCACTGTAGAT
PBoli 749	5’- phosphate –ATCTACAGTGAGTCGTACGC-3’-biotin
T3	5’- ATTAACCCTCACTAAAGGGA
T7	5’- TAATACGACTCACTATAGGG
XWP9	5’- AGTGTATTTTTTCGCCATTTACTCG
XWP10	5’- TAGGCGGATCGTGAAGTTAA

III.3.5 G overhang Capture Assay

Telomeres were cloned from *S. pombe* genomic DNA using the G overhang capture assay [40]. Oligonucleotides PBoli 733 and PBoli 749 (Table 3.2) were annealed into a partial duplex (0.5 pmol) by incubating at 95°C followed by gradual cooling to room temperature in 20mM Tris-acetate, pH 7.5, 50mM sodium chloride, and 2mM magnesium chloride. The partial duplex (500 fmol) was ligated to *S. pombe* genomic DNA with Quick Ligase and buffer supplied by the manufacturer (New England Biolabs) for fifteen minutes at room temperature. The ligase was heat inactivated at 65°C for twenty minutes before adding EcoRI (40U) and supplied buffer (New England Biolabs) to the reaction. Restriction digest was carried out for three hours at 37°C. Biotinylated products consisting of terminal DNA fragments were purified with magnetic streptavidin beads (Dynal) and unbound genomic DNA was removed with two washes in 10mM Tris-HCl, pH 8.0, 1mM EDTA, and 0.3M NaCl followed by two washes in 10mM Tris-HCl, pH 8.0, 1mM EDTA. Purified chromosome end fragments were amplified by PCR with PBoli 434 and PBoli 745 (Table 3.2). Amplification products were cloned into the pCR4-blunt-TOPO vector (Invitrogen) for analysis of individual telomeric sequences.

III.3.6 Telomere Sequence Analysis

Subtelomeric element sequence upstream of the noncanonical wild type repeats 5'-gggttacaaggttacgtggttacac-3' and plasmid sequence downstream of and including 5'-atctacagtgagtcgtacgc-3' were trimmed from raw data files of individual telomeric sequences from each strain. Trimming was performed with the "Trim Selected Fragments Ends For Vector Contamination…" function in ContigExpress and polylinkers corresponding to the sequence immediately upstream of the noncanonical wild type repeats (GTAAGGCGAGGCTGCG) and the ATCTACAGTGAGTCGTACGC sequence. Trimmed sequences were exported in FASTA format for further analysis. Symbol plots were generated to highlight repeats in telomeres of each strain. In the symbol plot the noncanonical wild type repeats were represented by a single triangle allowing up to three mismatches. A small set of telomeric sequences (9 of 2026) were severely shortened and missing part of the subtelomeric element. These sequences were analyzed separately.

III.3.7 End Fusions PCR

Junctions between fused chromosome ends were detected using a modified PCR from Wang et al. 2008 [99]. PCR reactions (25μl) consisted of 10X ThermoPol Buffer (New England Biolabs), 200μM dNTPs, 0.1μM each of primers BLoli 3400 and BLoli 3405 (Table 3.2), 0.5μM each of primers BLoli 1256 and BLoli 1353 (Table 3.2), 1.25 U Taq polymerase (New England Biolabs), and 1ng of genomic DNA. The cycling parameters were as follows: 94°C for 30 seconds, 32 cycles of 94°C for 10 seconds then 55°C for 30 seconds and 68°C for 3 minutes, with a final extension at 68°C for 10 minutes. DNA products were loaded onto a 1% agarose 0.25μg/ml ethidium bromide gel and imaged on the Typhoon 8600 scanner.

III.4 Results

III.4.1 Telomere length and growth phenotypes of 3'-CCAAU**N**U-5' template variants

The 3'-CCAAU sequence accounts for five of the six nucleotides in the wild type template that correspond to the core nucleotides found in all wild type repeats. TER1 G235 is the sixth nucleotide in this core. To determine the contribution of this position to telomere length and cell growth, strains were generated by transforming a *ter1*::*ura4* strain with a plasmid expressing wild type TER1 (3'-CCAAUGU-5') or one of the alternative TER1 templates (3'-CCAAU**U**U-5', 3'-CCAAU**A**U-5', and 3'-CCAAU**C**U-5') (Figure 3.1A). When grown independently on selective plates for 3 days, all of the alternative template strains grew similarly to wild type (Figure 3.1B). This finding supports our previous observation that the absence of a 3'-CCAAU-5' core is the major determinant of a growth defect. After ~75 generations, 3'-CCAAU**U**U-5' and 3'-CCAAU**A**U-5' yielded a range of telomere lengths closely resembling wild type (Figure 3.1C) with averages within 10nt from the wild type average of 146nt (Figure 3.1D). In contrast, 3'-CCAAU**C**U-5'demonstrated a marked increase in telomere length with an average length of 205nt (Figure 3.1.C, D). Together, these results demonstrate that beyond the 3'-CCAAU-5' core, the sixth core nucleotide of the TER1 template can be mutated and retain wild type growth for at least three days after plating. When the sixth nucleotide is a cytosine, telomeres are elongated without causing a growth defect.

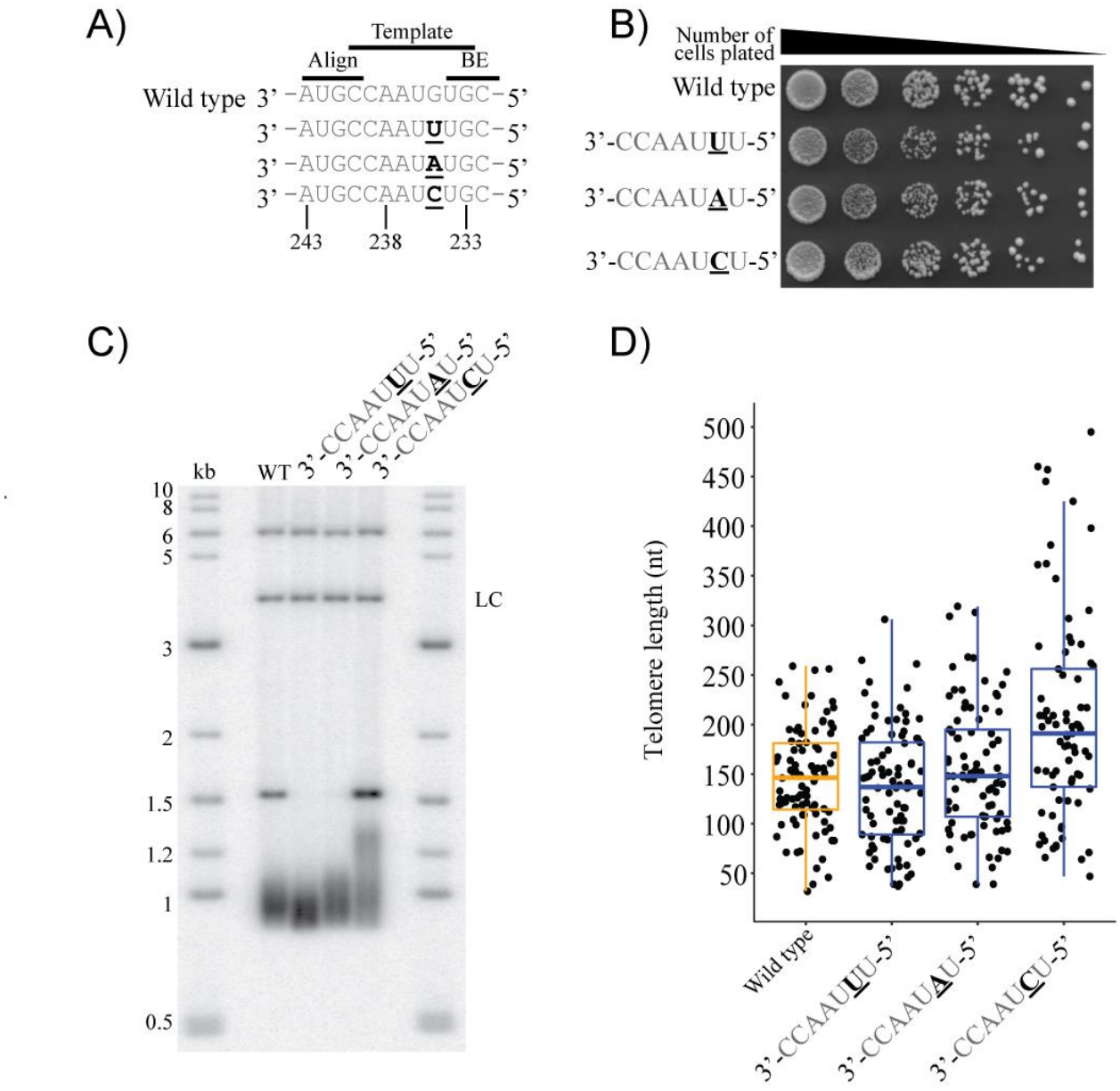


Figure 3.1 The 3’-CCAAU<u>C</u>U-5’ template confers increased telomere length. (A) TER1 template sequence of plasmid strains, mutation bolded and underlined. (B) Spotting assay with 1:5 serial dilutions starting from $5x10^4$ cells and plated on YEA + 100µg/ml Geneticin, incubated at 32°C, and grown for 3 days. (C) Telomere length analysis by southern blot hybridized with a probe specific for the subtelomeric element sequence, wild type (wt). A probe specific to rad16 was used as the loading control (LC). (D) Telomere length analysis by counting nucleotides of cloned telomeres. Absolute numbers are shorter than southern blot analysis because the subtelomeric element sequence was trimmed from raw data files.

III.4.2 The 3'-CCAAU**C**U-5' template causes a shift in the putative alignment region

To determine if the 3'-CCAAU**C**U-5' TER1 template was used to template the addition of repeats at the chromosome end, telomeres of this strain were cloned and sequenced. The most abundant motifs in the telomere were 5'-GGTTAGA-3', 5'-GTACGGTTAGA-3', and 5'-GGTTAGAGTAC-3'. These motifs formed repetitive sequences with 5'-GGTTAGA-3' making up doublet or triplet repeats (Figure 3.2.A and Figure 3.3). When 5'-GTAC-3' was included in the repeat motif, two to four consecutive repeats were observed. Because 5'-GTAC-3' is not always followed by 5'-GGTTAGA-3', we observed a slightly greater number of consecutive repeats with the 5'-GGTTAGAGTAC-3' motif than the 5'-GTACGGTTAGA-3' motif. The spacer sequences between 5'-GGTTAGA-3' motifs were determined. Just as expected, 5'-GTAC-3' was the most frequent spacer sequence and was often included among longer spacers (Figure 3.2.B). Analysis of different scenarios for the alignment of the 5'-GGTTAGA-3' repeat and 5'-GTAC-3' spacer to the alternative template TER1 revealed that the telomeric 5'-GTAC-3' is most likely reverse transcribed from G_{241} to C_{244} of TER1 (Figure 3.2C). The reverse transcription of $G_{241} - C_{244}$ would require the last three nucleotides of the 5'-GGTTAGA-3' repeat to align with 3'-UCU_{245} of TER1. If the wild type alignment register, 3'-AUG_{241}, is used instead, then the other observed spacer sequences such as 5'-GGTTA-3' and 5'-GTTAC-3' would be incorporated (Figure 3.2D). In conclusion, the 3'-CCAAU**C**U-5' variant incorporates templated alternative repeats into telomeres. Use of this alternative template involves alignment to an unmutated sequence that is shifted four nucleotides downstream from the wild type alignment region. In wild type, telomeric DNA was shown to translocate, or align, in three different registers of the RNA alignment-template [40]. Contrastingly, telomere repeats reverse transcribed from the 3'-CCAAU**C**U-5' template can switch between a 3'-UCU_{245} alignment and

a 3’-AUG_{241}, signifying surprising flexibility in the telomeric DNA-telomerase RNA-catalytic subunit interaction.

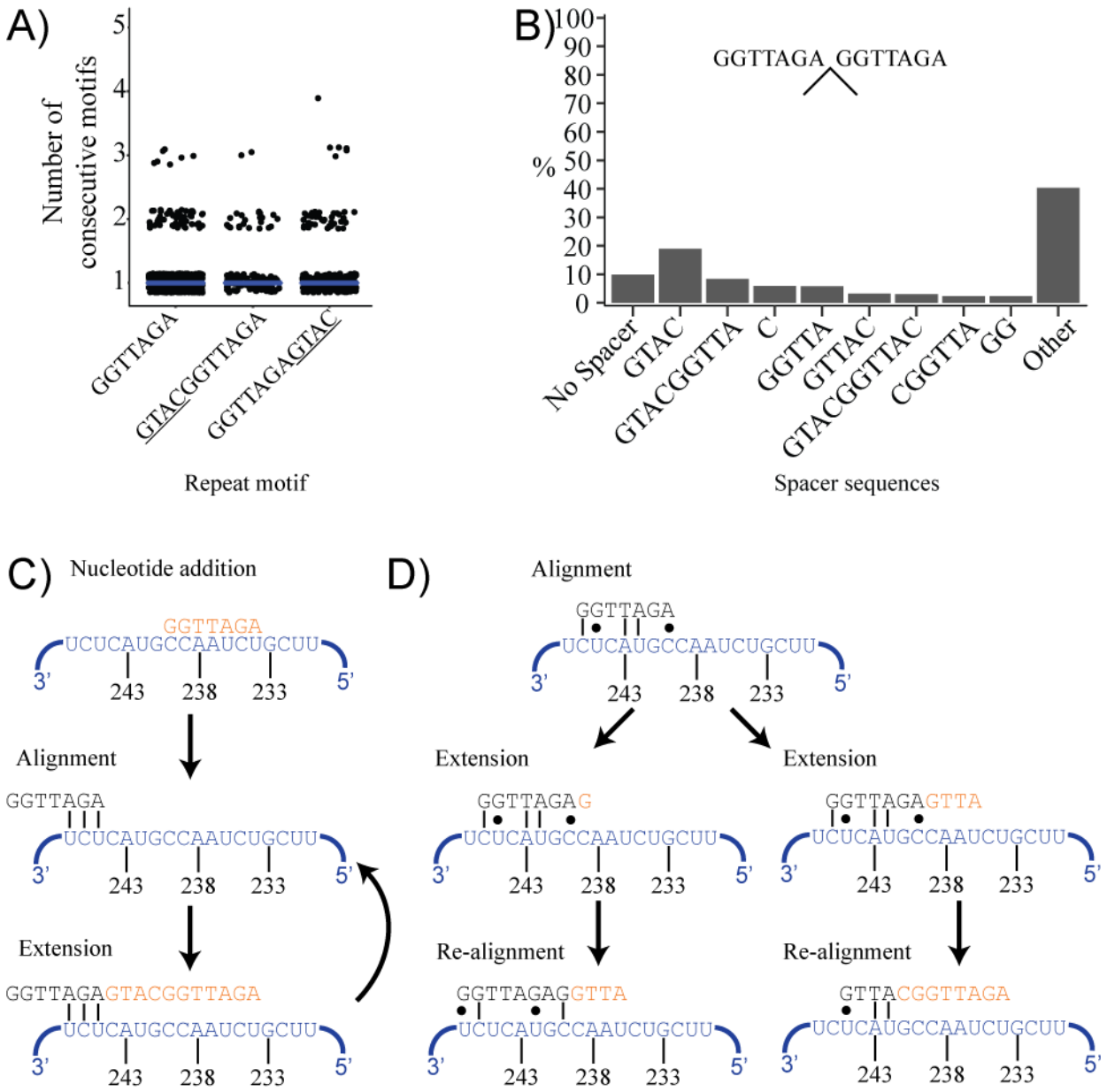


Figure 3.2 Telomeric 5'-GTAC-3' sequence corresponds to C_{244} to G_{241} of TER1. Sequences of 77 telomeres (15,763nts combined). (A) Sequencing data were analyzed for mutant motifs that generated the highest number of consecutive repeats. The three motifs yielding the highest consecutive repeat numbers are shown. (B) Frequency of sequences interrupting GGTTAGA repeats. The "other" category includes spacer sequences with a frequency < 2%. (C, D) Proposed alignments of 3' telomeric overhang (black) and TER1 sequence (blue) to generate a templated repeat (orange). Vertical lines denote canonical Watson-Crick base pairs and a single dot represents noncanonical base pairing between telomere and TER1. (C) Alignment to 3'-UCU-5'of TER1 and corresponding inclusion of 5'-GTAC-3' in the telomere. (D) Scenario of alternative alignments that would result in the inclusion of experimentally observed spacer sequences.

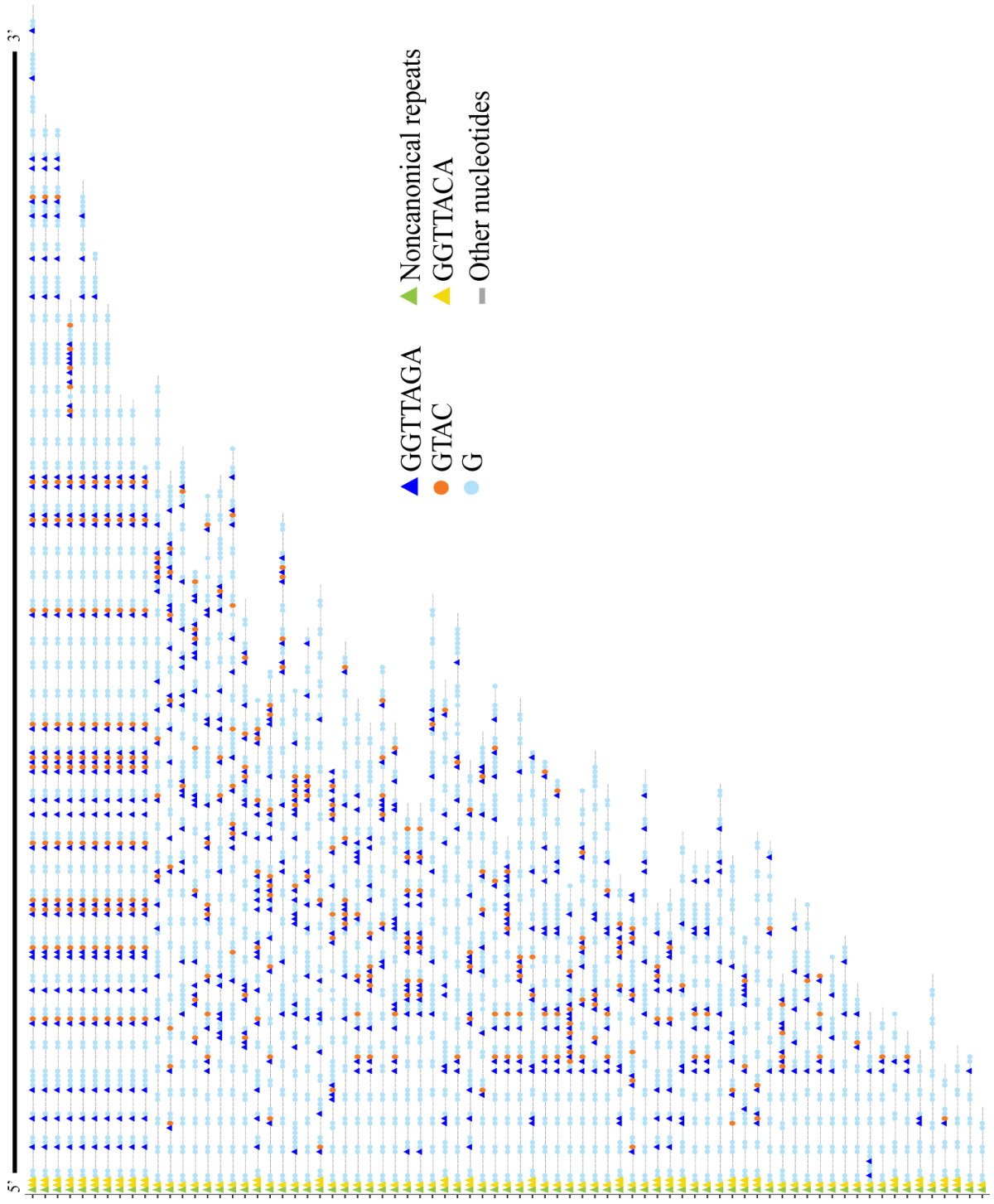


Figure 3.3 The 3'-CCAAUCU-5' strain templates a 5'-GGTTAGA-3' repeat. Alignment of all cloned telomeres. Remnants of the wild type telomere including noncanonical repeats (green triangle), 5'-GGTTACA-3' (gold triangle), and 5'-GGTTAC-3' are clustered toward the proximal end of the telomere. Mutant repeats (blue triangle) are often accompanied by a 5'-GTAC-3' sequence (orange circle).

III.4.3 Telomere length and growth phenotypes of 3'-CCAAUG**N**-5' template variants

The seventh nucleotide in the template overlaps with the boundary element in wild type TER1. We investigated the impact of point mutations at this position using the following plasmid strains named for their template sequence: 3'-CCAAUG**G**-5', 3'-CCAAUG**A**-5', and 3'-CCAAUG**C**-5' (Figure 3.4.A). There was no appreciable growth difference between these strains and wild type when grown independently on selective plates (Figure 3.4.B). After ~75 generations, 3'-CCAAUG**G**-5' and 3'-CCAAUG**A**-5' showed greatly reduced telomere lengths (Figure 3.4C). The average length of cloned wild type telomeres was 146nt compared to 55nt and 77nt for 3'-CCAAUG**G**-5' and 3'-CCAAUG**A**-5', respectively (Figure 3.4D). Telomere length was preserved in 3'-CCAAUG**C**-5' with an average of 152nt among cloned telomeres (Figure 3.4C, D).

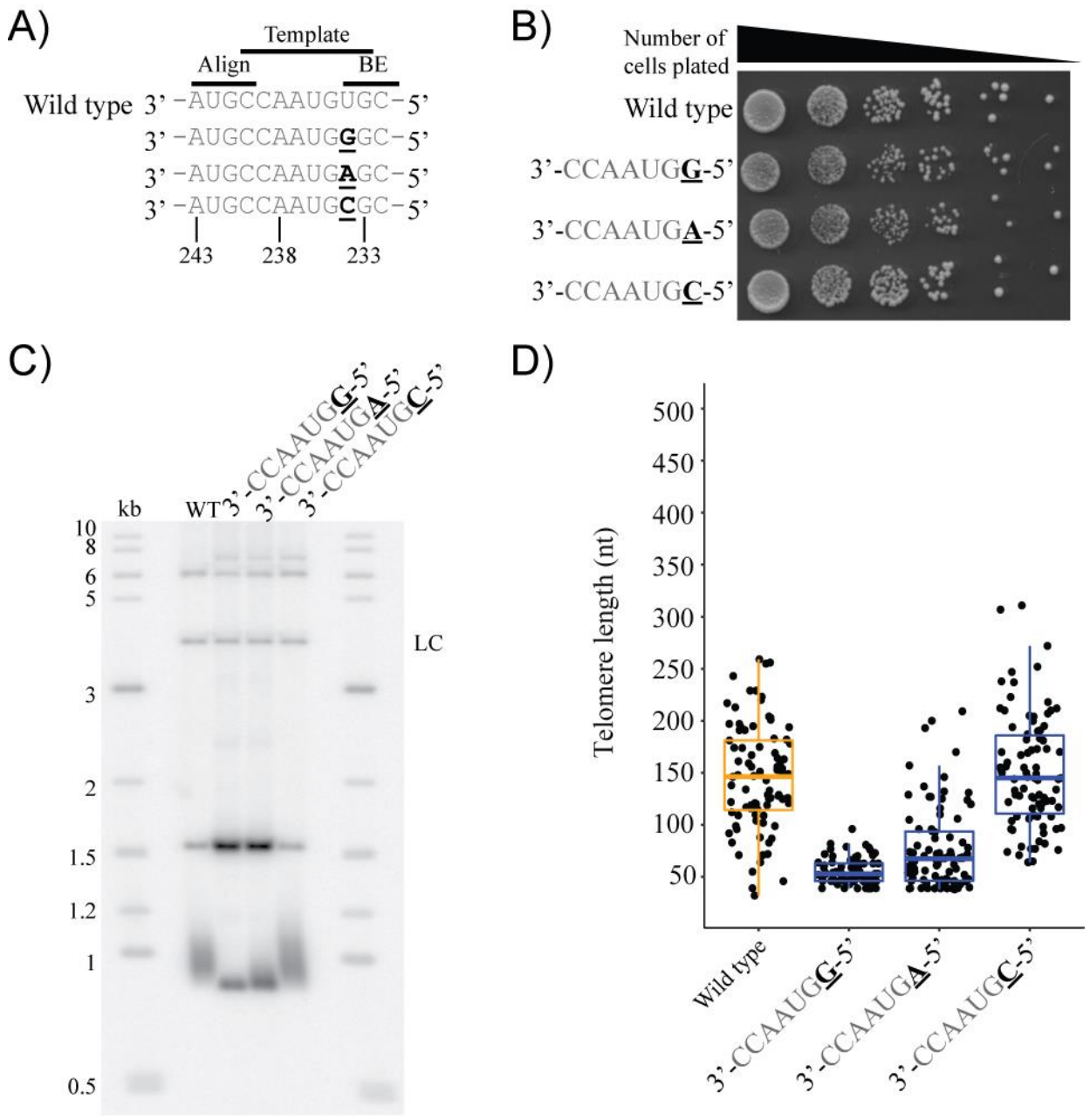


Figure 3.4 The 3’-CCAAUGC-5’ template conserves wild type telomere length. (A) List of plasmid strains with corresponding TER1 template sequence and mutation (bolded and underlined). (B) Spotting assay with 1:5 serial dilutions starting from $5x10^4$ cells on YEA 100μg/ml Geneticin plates and incubated at 32°C for 3 days. (C) Southern blot probed for subtelomeric element sequence, wild type (wt). A rad16 specific probe was used as the loading control (LC). (D) Telomere length in nucleotides (nt) of cloned telomeres. Reported values are nucleotides counts after trimming the subtelomeric element sequence.

III.4.4 Loss of repeat heterogeneity in the 3'-CCAAUG**C**-5' TER1 variant strain

The wild type plasmid strain demonstrated the same telomeric repeat patterns as previously described for endogenous *ter1*$^{+}$ strains [45, 46, 60, 61, 226, 227] . The first and second most common motifs of all wild type repeats is 5'-GGTTACA-3'and 5'-GGTTAC-3', respectively [40]. Similarly, a small sampling of telomeric sequences from the wild type plasmid strain confirmed 5'-GGTTACAGG-3' as the most abundant repeat motif followed by 5'-GGTTACGG-3' (Figure 3.5). As expected, 5'-GGTTACA-3' was only included in repeats of 2-4 motifs (Figure 3.6B). The number and sequence of spacer nucleotides between 5'-GGTTAC-3' motifs were comparable to those previously reported by Trujillo *et al* [61]. Consistent with 5'-GGTTACA-3' being the most common repeat, a single adenine was the most frequent sequence between 5'-GGTTAC-3' motifs, accounting for 40% of all intervening sequences (Figure 3.6C, D). Beyond that, spacers, 2-5nt in length and incorporating a variable number of guanines or an adenine-cytosine sequences, were the most frequent (Figure 3.6C, D).

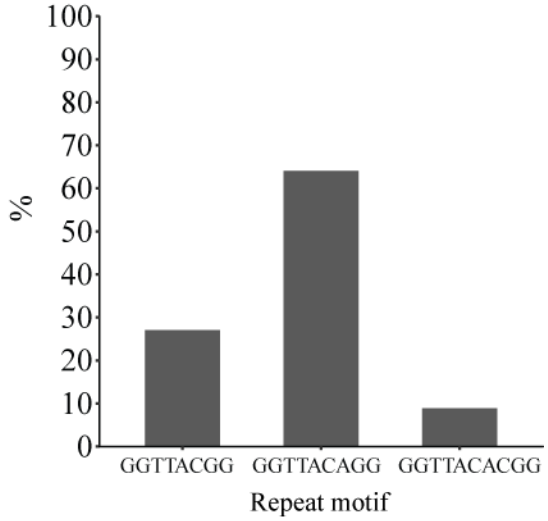


Figure 3.5 5'-GGTTACA-3' motif is the most common repeat in the wild type plasmid strain. Frequency of each repeat motif was determined in a set of 88 telomeres (12,811nts combined).

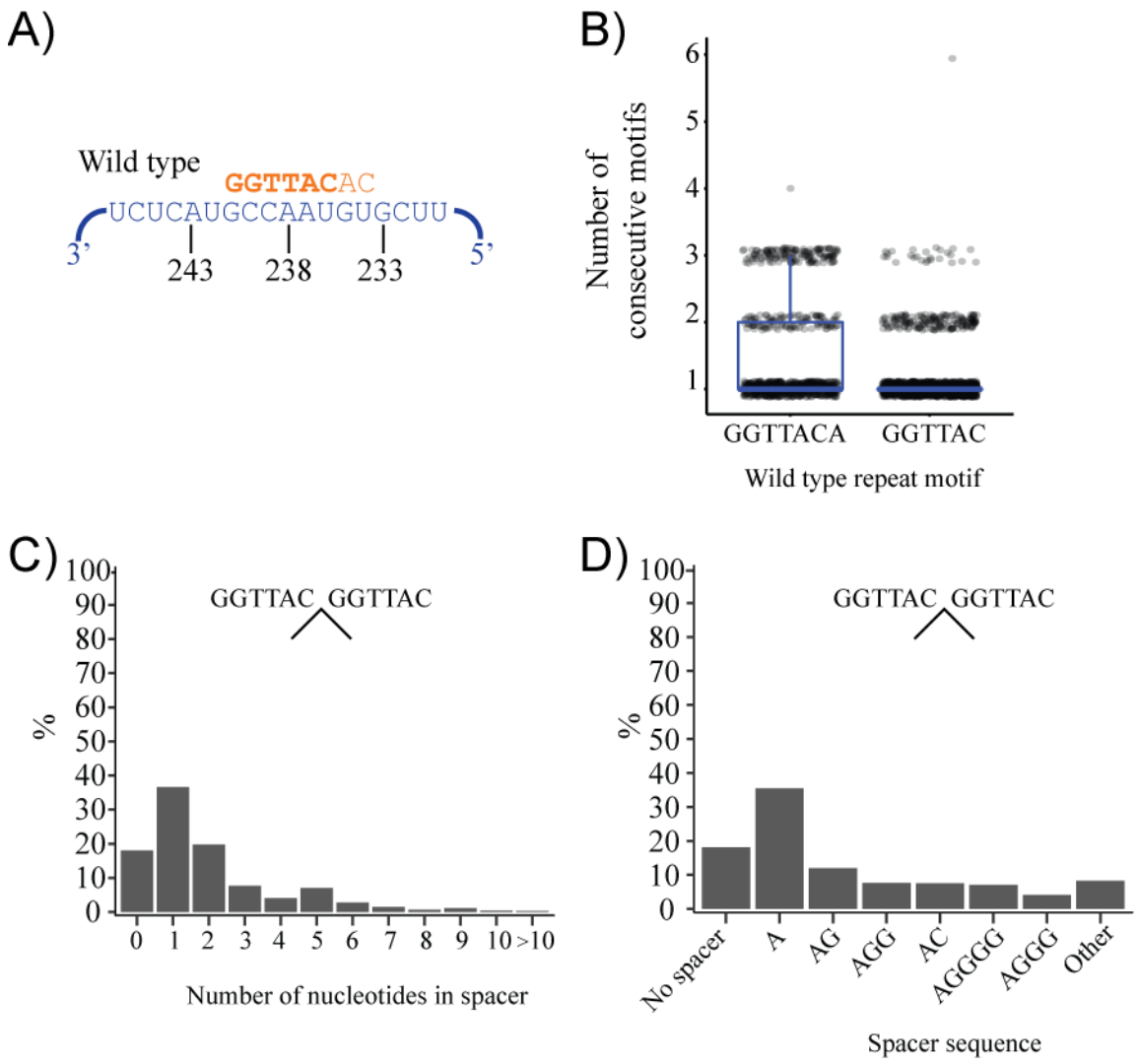


Figure 3.6 Wild type plasmid strain demonstrates heterogeneous telomere phenotype. Sequences of 88 telomeres (12,811nts combined). (A) Alignment of wild type TER1 (blue) to telomeric 3' overhang (black) and variations of the newly synthesized repeat (bold and light orange). Vertical lines represent Watson-Crick base pairs and a single point denotes noncanonical base pairs. (B) Consecutive repeat motifs were counted independently within a single telomere. Sequencing data were processed into GGTTAC repeats and analyzed for the length of sequences between two repeats (C) and the sequence of nucleotides between two repeats; spacer sequences with a frequency < 2% are combined in the "other" category (D).

Surprisingly, whereas wild type generated highly heterogeneous repeats (Figure 3.7), 3'-CCAAUG**C**-5' generated nearly perfect repeats (Figure 3.8). Accordingly, slightly more than 70% of repeats in 3'-CCAAUG**C**-5' did not have a spacer. Because the 3'-CCAAUG**C**-5'

variant template was introduced into cells after the loss of endogenous *ter1+* locus, the variant-template telomerase must have aligned to wild type repeats to elongate the 3' overhang. Therefore, 2-3 wild type 5'-GGTTACA-3' repeats were observed toward the proximal telomere with an increasing likelihood of variant-template telomeric repeats toward the distal telomere. Interestingly, we frequently noted a G-rich transition zone between the last 5'-GGTTACA-3' and the first repeat of a 5'-GGTTAC-3' stretch. This finding indicates a stuttering of telomerase in the transition from a wild type repeat-variant alignment/template interaction to a variant repeat-variant alignment/template interaction. Of note, this template strain demonstrates that a single nucleotide change can confer a nearly perfect repeat phenotype in lieu of degenerate repeats. In summary, the study of point mutants of the single nucleotide overlap between the wild type template and boundary element demonstrated that the 3'-CCAAUG**C**-5' mutation can change telomeric repeats from being highly heterogeneous to nearly perfect following a stuttering-mediated transition zone.

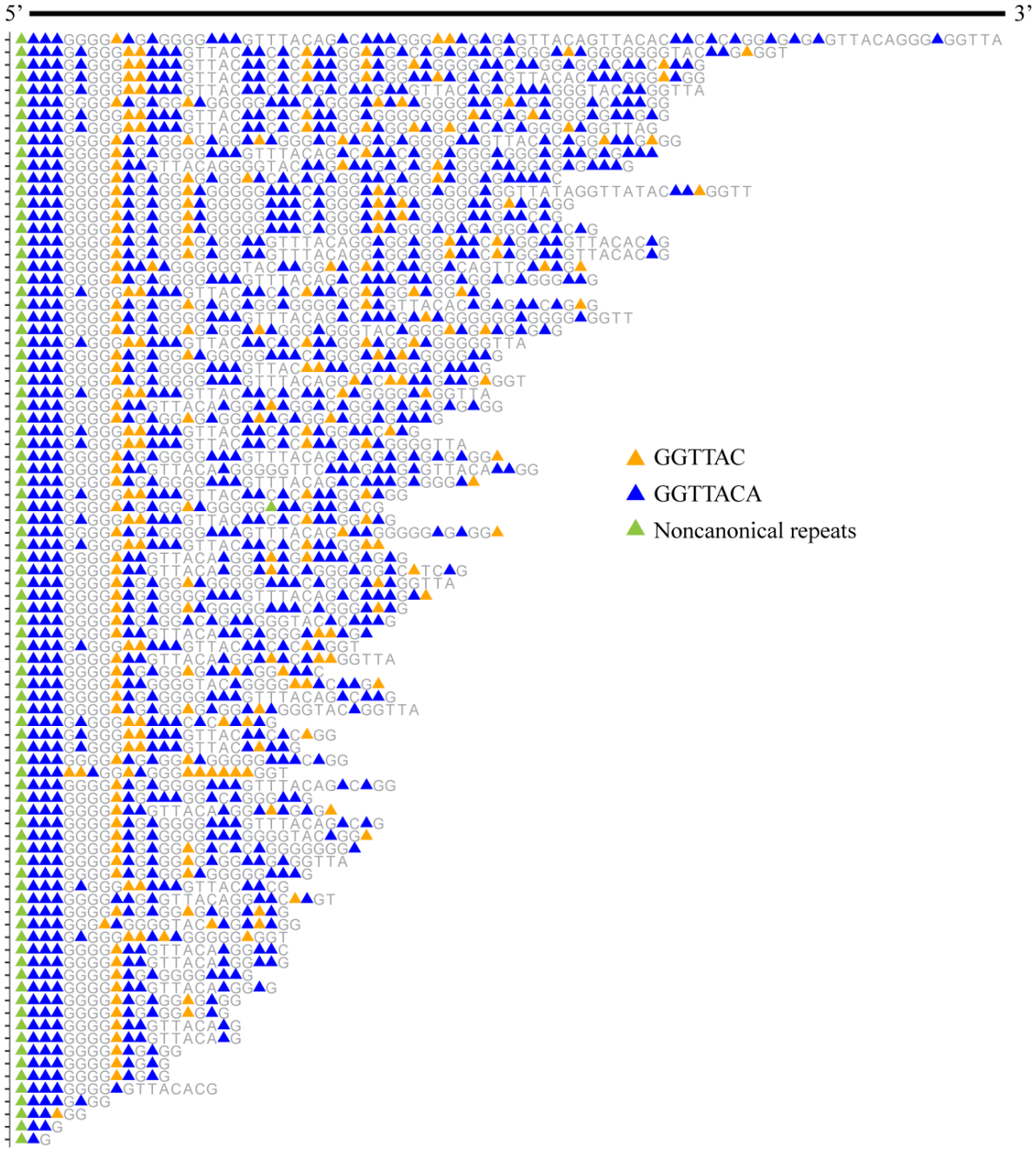


Figure 3.7 Wild type heterogeneous telomeres are comprised of $GG_{1\text{-}6}TTACA_{0\text{-}1}C_{0\text{-}1}$ repeats. Alignment of all cloned telomeres from the wild type plasmid strain. All wild type telomeres begin with invariant noncanonical repeats (green triangle). Two of the most frequent repeats are highlighted, 5'-GGTTACA-3' (blue triangle) and 5'-GGTTAC-3' (orange triangle).

Figure 3.8 The 3'-CCAAUG<u>C</u>-5' strain templates nearly perfect 5'-GGTTAC-3' repeats. Alignment of all cloned telomeres from the 3'-CCAAUG**<u>C</u>**-5' plasmid strain. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' (gold triangle) are clustered toward the proximal end of the telomere. Mutant repeats (blue triangle) follow a transition zone made mostly of single guanines.

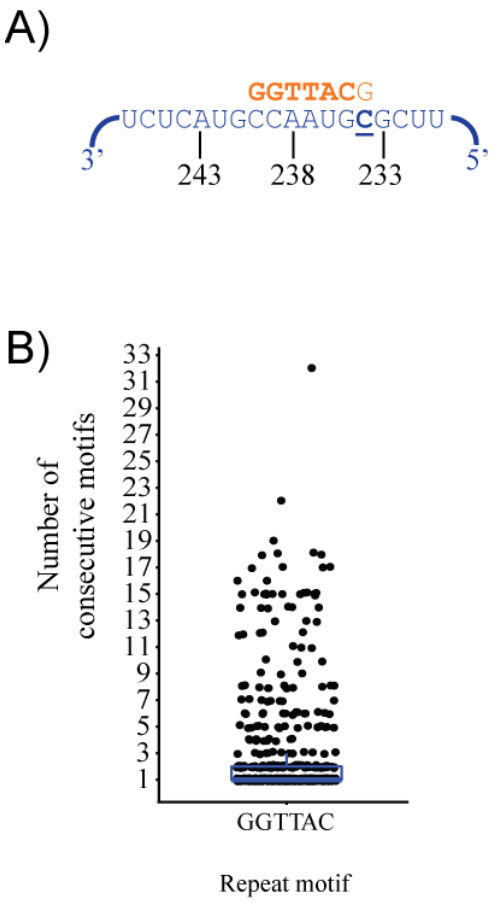


Figure 3.9 Consecutive repeat number in 3'-CCAAUG<u>C</u>-5' template strain. Sequences of 86 telomeres (13,002nts combined). (A) Alignment of 3'-CCAAUG**<u>C</u>**-5' TER1 (blue; mutation bolded and underlined) to variations of the newly synthesized repeat (bold and light orange). (B) Sequencing data were analyzed for motifs that generated the highest number of consecutive repeats throughout the telomere. Stretches of motifs were counted independently. Sequencing data were processed into GGTTAC repeats and analyzed for the length of sequences between two repeats.

III.4.5 Determination of telomeric repeat sequence in mutant template strains

A key difference between 3'-CCAAUG**<u>C</u>**-5' and wild type telomeres is that the former most frequently incorporates a repeat ending in 5'-TAC-3' while the most frequent wild type repeat ends in 5'-ACA-3'. When a repeat ending in 5'-ACA-3' is incorporated at the chromosome end, wild type telomerase makes use of a 3'-UGC-5' alignment and C_{240} +1 register to copy a single guanine from C_{239} (Figure 1.5). Once the single guanine is incorporated, telomerase translocates to the 3'-AUG-5' alignment region and continues to copy single guanines, or "stutter", until the full 5'-GGTTACA-3' repeat is completed. However, when the wild type repeat ends in 5'-TAC-3', the TER1 3'-AUG-5' alignment region can be used in the

immediate next step. In 3'-CCAAUG**C**-5', the telomeric 5'-TAC-3' can be reverse transcribed from the TER1 3'-AUG_{235} (Figure 3.9A). Therefore, I hypothesized that telomeres from this strain included nearly perfect repeats because most repeats ended in 5'-TAC-3' which could then perfectly base pair with the 3'-AUG_{241}-5' in the TER1 alignment region during translocation. This hypothesis would be disproved by a strain with most repeats ending in 5'-TAC-3' but a lower number of consecutive repeats than wild type or by a strain where most repeats do not end in 5'-TAC-'3' and, yet, have a greater number of consecutive repeats compared to wild type.

To test this hypothesis, I generated plasmid strains with templates harboring the same two variable nucleotides beyond the 3'-CCAAU-5' core, but in different locations. I then cloned and sequenced their telomeres. The location of the variable nucleotides could result in either the removal of 3'-AUG-5' from the template or its inclusion into the template, which would cause a 5'-TAC-3' in the telomeric repeat. The repeats of these strains were defined by a "kmer" analysis of nucleotide patterns of length "k", TweenMotif [61], counts of consecutive repeat motifs, a study of spacer sequences, and alignment scenarios between template and telomere. This analysis was complicated by the nature of these variant telomeres that contained a wild type proximal portion and a mutant distal portion. Still, we were able to define repeats for three sets of templates described below.

Example 1: Variable nucleotides 3'-GC-5'

The first set of templates shared the 3'-GC-5' variable nucleotides and included the following templates a) 3'-CCAAU**GC**-5', b) 3'-**G**CCAAU**C**-5', and c) 3'-**GC**CCAAU-5'. The 3'-**G**CCAAU**C**-5' template strain generated highly heterogeneous repeats best represented by 5'-$C_{0\text{-}2}G_{1\text{-}3}TT_{1\text{-}6}A_{1\text{-}5}$-3' (Figure 3.10A). Interestingly, this strain demonstrates for the first time that

the position of two variable nucleotides in relationship to the 3'-CCAAU-5' core can affect telomerase translocation. In addition to increased repeat heterogeneity, the telomeres of this strain exhibit a thymidine (T)-stutter pattern rather than the wild type guanine (G)-stutter pattern. This phenotype suggests slippage from A_{236} to A_{237} to generate 2-6 consecutive T's in the telomere (Figure 3.10B). The 3'-**GC**CCAAU-5' template strain generated more stereotyped repeats defined as 5'-CGGGTTAC-3' (Figure 3.11, Figure 3.12).

With telomere sequences in hand, we next tested the hypothesis that repeats ending in 5'-TAC-3' resulted in an increased consecutive repeat number. In support of the hypothesis, the two templates with a 3'-AUG-5', 3'-CCAAU**GC**-5' and 3'-**GC**CCAAU-5', and repeats ending in 5'-TAC-3', also demonstrated an increase in consecutive repeat number compared to the third template, 3'-**G**CCAAU**C**-5', that did not have a 3'-AUG-5' in the template (Figure 3.13).

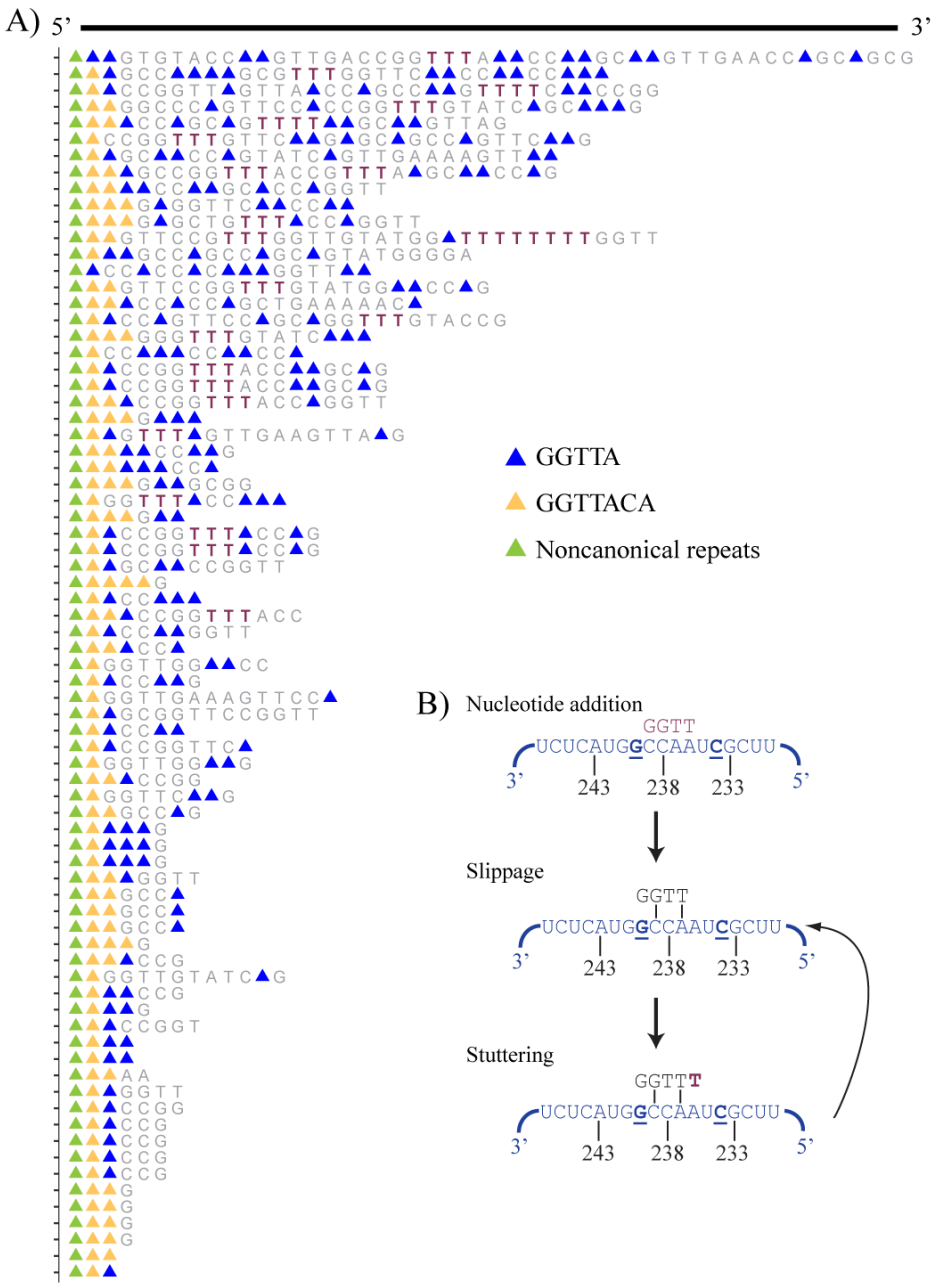


Figure 3.10 3'-GCCAAUC-5' template generates heterogeneous repeats. A) Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-GCCAAUC-5'. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3'repeats (gold triangle) are clustered toward the proximal end of the telomere. The most abundant sequence 5'-GGTTA-3' is highlighted by a blue triangle. Stretches of 3 or more thymidines are shown in purple. B) Translocation cycle that generates consecutive thymidines in the degenerate repeat. Telomeric 3' overhang (black), TER1 (blue, mutant nucleotides of template are bolded), newly synthesized nucleotides (purple).

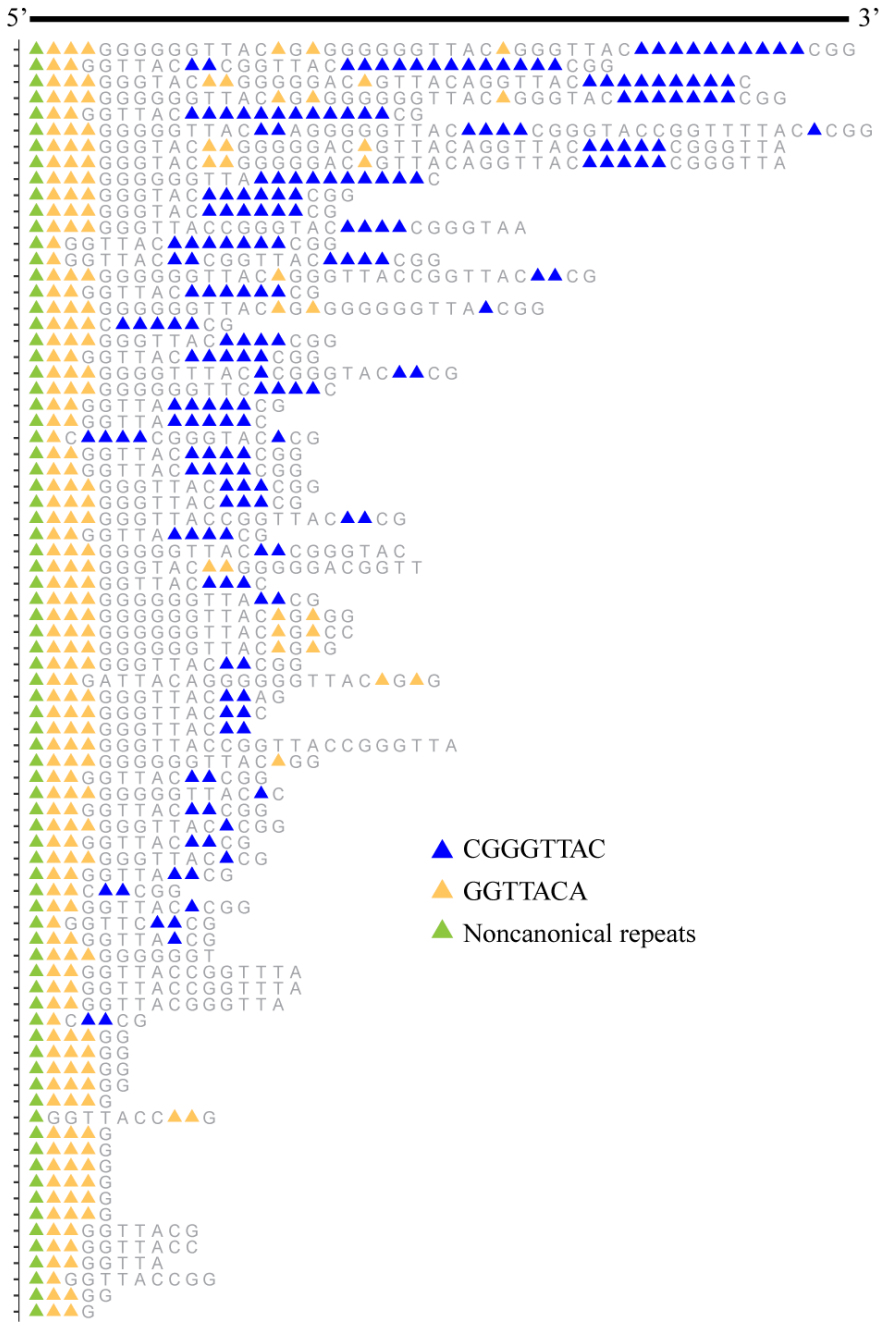


Figure 3.11 5'-CGGGTTAC-3' is the most common and consecutive repeat generated from the template 3'-<u>GC</u>CCAAU-3'. Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-<u>**GC**</u>CCAAU-5'. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3'repeats (gold triangle) are clustered toward proximal end of the telomere. The most frequent repeat 5'-CGGGTTAC-3' is highlighted by a blue triangle.

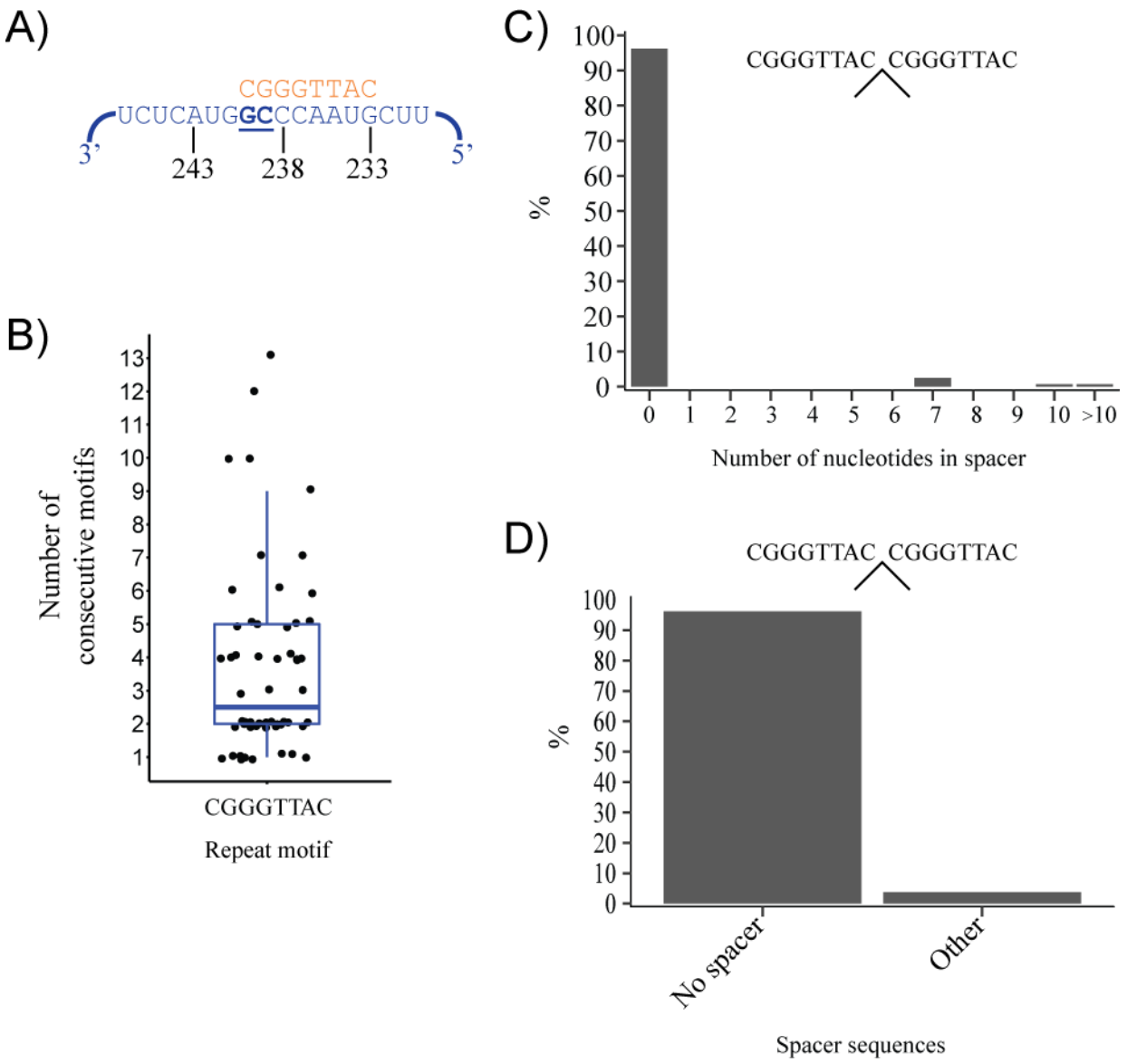


Figure 3.12 3'-<u>GC</u>CCAAU-5' template generates highly consecutive repeats. Sequences of 79 telomeres (5,998nts combined). (A) Alignment of 3'-**<u>GC</u>**CCAAU-5' TER1 (blue; mutation bolded and underlined) to the alternative repeat (orange). (B) Sequencing data were analyzed for the most frequent and repetitive motifs. Stretches of motifs were counted independently. Sequencing data were processed into CGGGTTAC repeats and analyzed for the length of sequences between two repeats (C) and the sequence of nucleotides between two repeats; spacer sequences with a frequency <2% are combined in the "other" category (D).

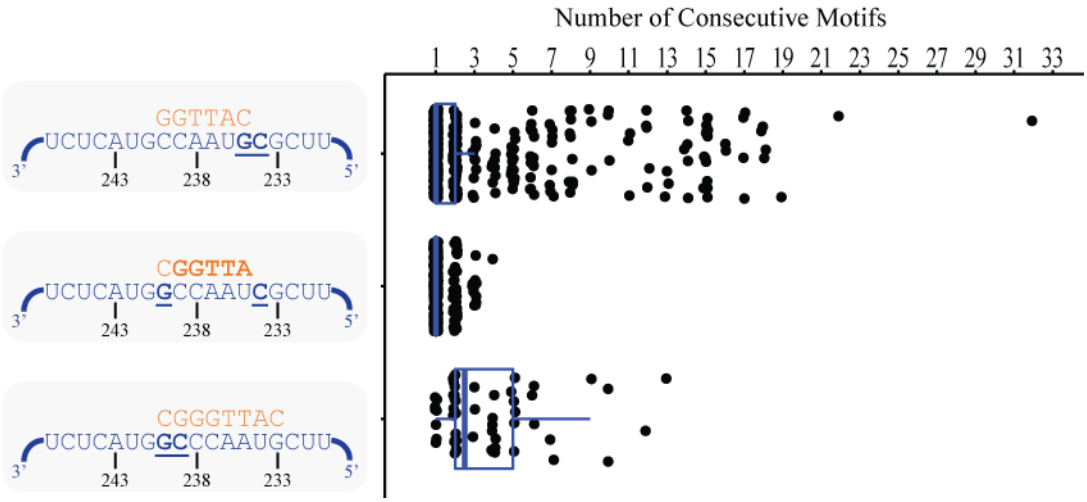


Figure 3.13 Example 1: Highly consecutive repeats end in 5'-TAC-3' templated by 3'-AUG-5' in the alternative TER1 template. The most frequent and repetitive motif (orange) of each strain was used to count the number of consecutive motifs within a stretch of repeats. Each stretch of repeats was counted independently in each telomere. TER1 (blue); variable nucleotides (bolded and underlined).

Example 2: Variable nucleotides 3'-AG-5'

We performed the same analysis for a second set of templates, which had the variable two nucleotides, 3'-AG-5'. This set included the 3'-**AG**CCAAU-5' strain, which incorporated stretches of eight-nucleotide 5'-TCGGTTAC-3' repeats into telomeres (Figure 3.14, Figure 3.15A). This strain was found to have a maximum of 29 consecutive repeats (Figure 3.15B), second only to the 3'-CCAAU**GC**-5' template strain (Figure 3.9B). More than 95% of repeats had no intervening sequences (Figure 3.15C, D). The 3'-**A**CCAAU**G**-5' template strain displayed 5'-TGGTTAC-3' repeats (Figure 3.16, Figure 3.17A). No spacers were present between 82% of repeats (Figure 3.17B). The only spacer sequence to account for >2% of spacers was 5'-CGGTTAC-3' which is generated by the nucleotide addition of an elongated, 5'-TGGTTACC-3' repeat followed by a translocation step (Figure 3.17C,D). Lastly, GGTTATCC

was the most common and the longest motif in telomeres from the template strain, 3'-CCAAU**AG**-5' (Figure 3.18). Two permutations of this motif 5'-GGTTATCC-3' or 5'-GTTATCCG-3' were indistinguishable based on the likelihood of reverse transcription from the template and the range of consecutive repeats (Figure 3.19 A, B). Overall, telomeres consisted of a heterogeneous sequence. An intervening sequence was not present between 49% of repeats (Figure 3.19C). Intervening sequences appear to be a partial repeat (Figure 3.19D) indicating either a premature translocation step during repeat synthesis or trimming of the 3'end prior to telomere elongation. The most frequent spacer is a single G nucleotide. It most likely results from stuttering at position C_{239} rather than reverse transcription of C_{232} because the latter would result in a more unstable alignment during the translocation step. The most frequent, longest spacer is 5'-GGTTATC-3'. Nucleotide addition that stops at G_{234} confers this sequence (Figure 3.19A). Then, 5'-GGTTATC-3' would translocate to align its last three nucleotides with 3'-AUG_{241}-5' of TER1 allowing for subsequent 5'-GGTTATCC-3' synthesis. The spacer sequences of this template strain liken it to the degenerative repeats found in wild type.

As expected the strains with 3'-AUG-5' in the template reverse transcribed repeats with 5'-TAC-3' and displayed a range of consecutive repeats that was increased above the template without 3'-AUG-5' in the template (Figure 3.20). The findings in this second example provide further support of the hypothesis. For reasons that remain unclear, the spread of consecutive repeats can vary by template, indicating a complex process for successive repeat synthesis that involves more than a matched 3'-AUG-5' sequence in the template and alignment regions.

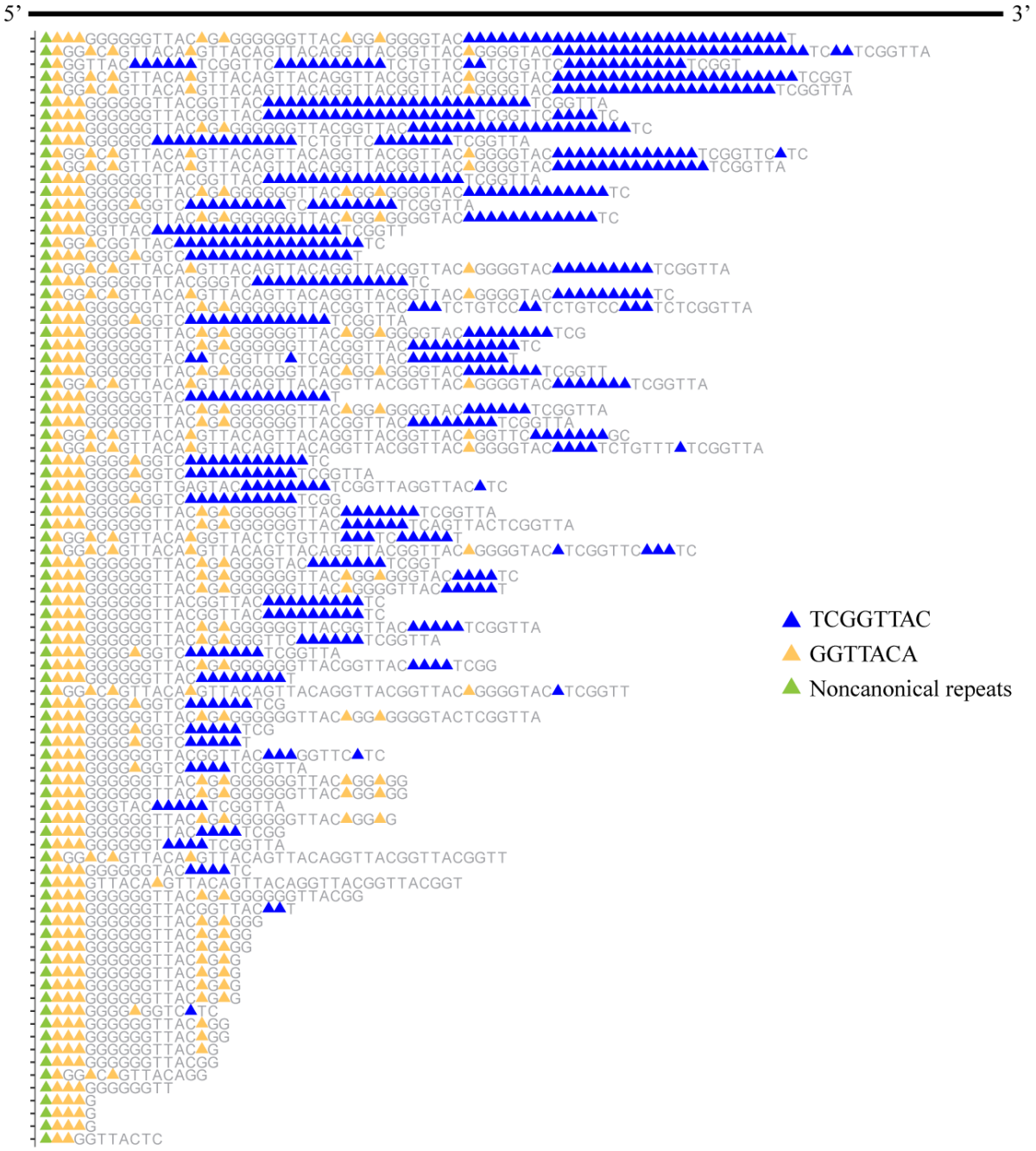


Figure 3.14 5'-TCGGTTAC-3' is the most common and consecutive repeat generated from the 3'-AGCCAAU-5' template. Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-AGCCAAU-5'. Remnants of the wild type telomere are highlighted including noncanonical repeats (green triangle) and 5'GGTTACA-3' repeats (gold triangle). The most frequent repeat 5'-TCGGTTAC-3' is denoted by a blue triangle.

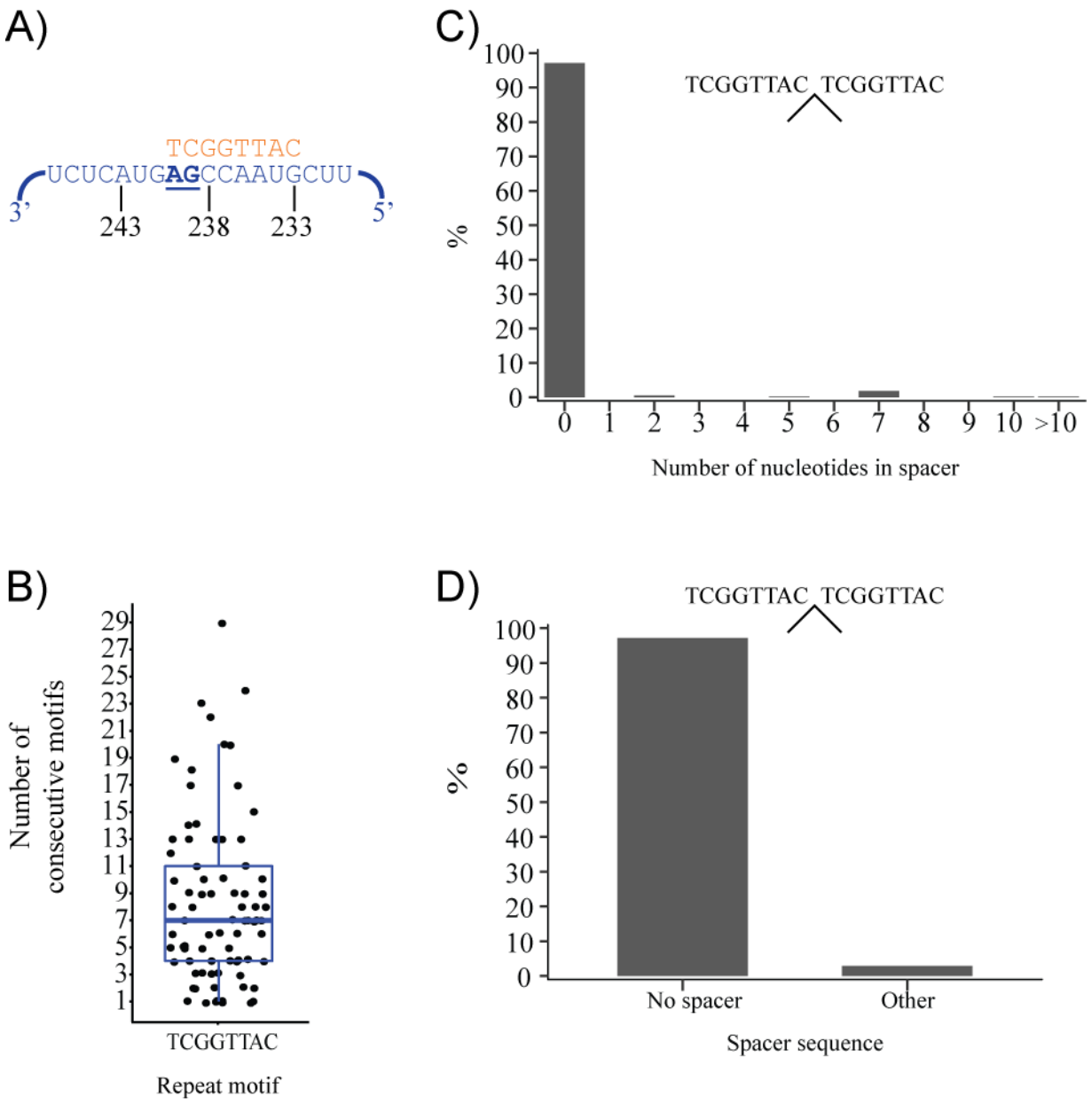


Figure 3.15 3'-<u>AG</u>CCAAU-5' template generates highly consecutive repeats. Sequences of 86 telomeres (13,002nts combined). (A) Alignment of 3'-**<u>AG</u>**CCAAU-5' TER1 (blue; mutation bolded and underlined) to the alternative repeat (orange). (B) The most frequent and repetitive motif is shown. Stretches of motifs were counted independently. Sequencing data were processed into TCGGTTAC repeats and analyzed for the length of sequences between two repeats (C) and the sequence of nucleotides between two repeats; spacer sequences with a frequency <2% are combined in the "other" category (D).

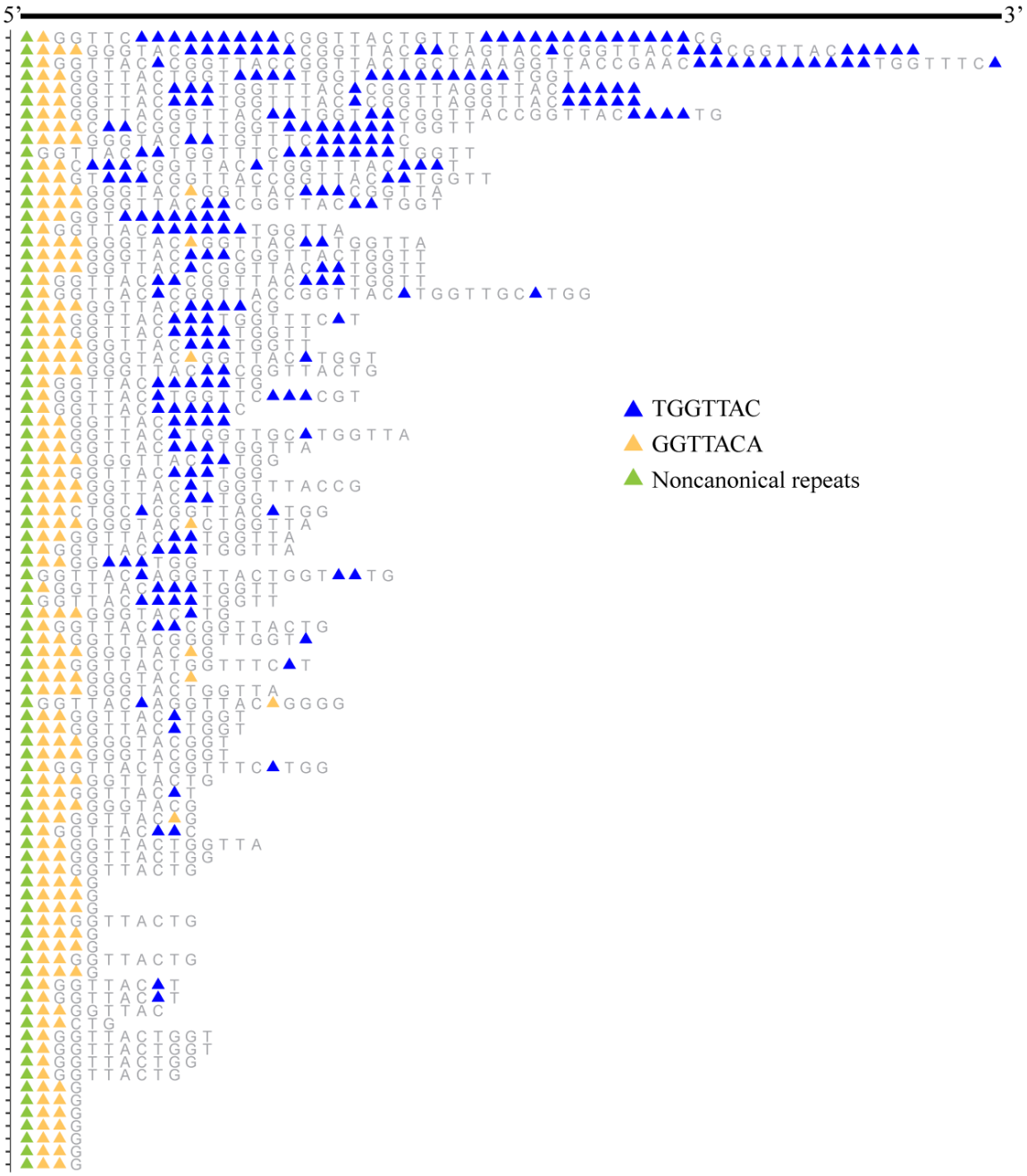


Figure 3.16 5'-TGGTTAC-3' is the most common and consecutive repeat generated from the template 3'-ACCAAUG-5'. Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-ACCAAUG-5'. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' repeats (gold triangle) are clustered toward the proximal end of the telomere. The most frequent repeat 5'-TGGTTAC-3'is highlighted by a blue triangle.

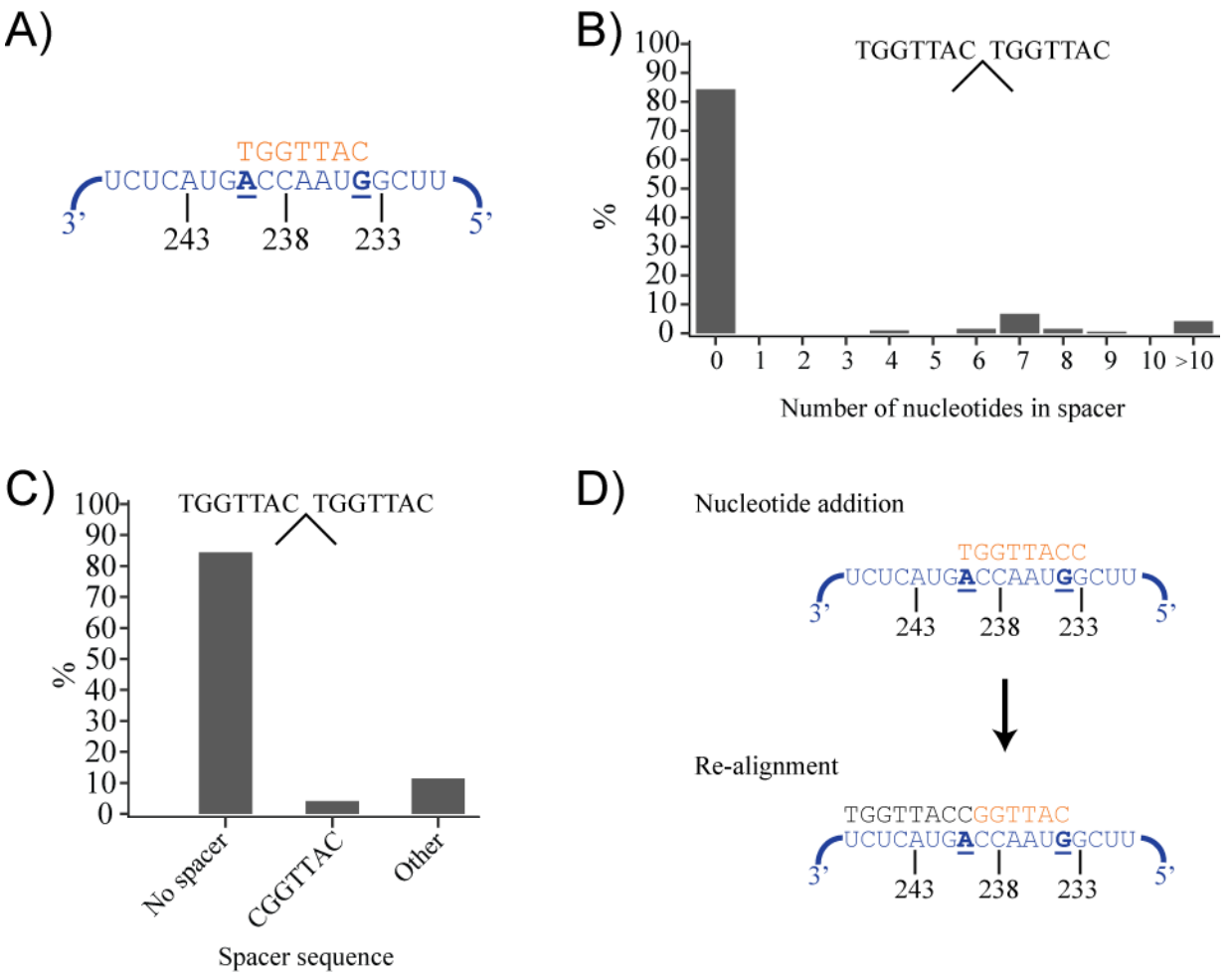


Figure 3.17 Synthesis of 5'-TGGTTACC-3' followed by a translocation step results in 5'-CGGTTAC-3' spacers. Sequences of 89 telomeres (6,216nts combined). (A) Alignment of 3'-**A**CCAAU**G**-5' TER1 (blue; mutation bolded and underlined) to the alternative repeat (orange). Sequencing data were processed into TGGTTAC repeats and analyzed for the length of sequences between two repeats (B) and the sequence of nucleotides between two repeats ; spacer sequences with a frequency < 2% are combined in the "other" category (C). (D) Nucleotide addition and realignment to generate the 5'-CGGTTAC-3' spacer sequence. Telomeric DNA of 3' overhang shown in black.

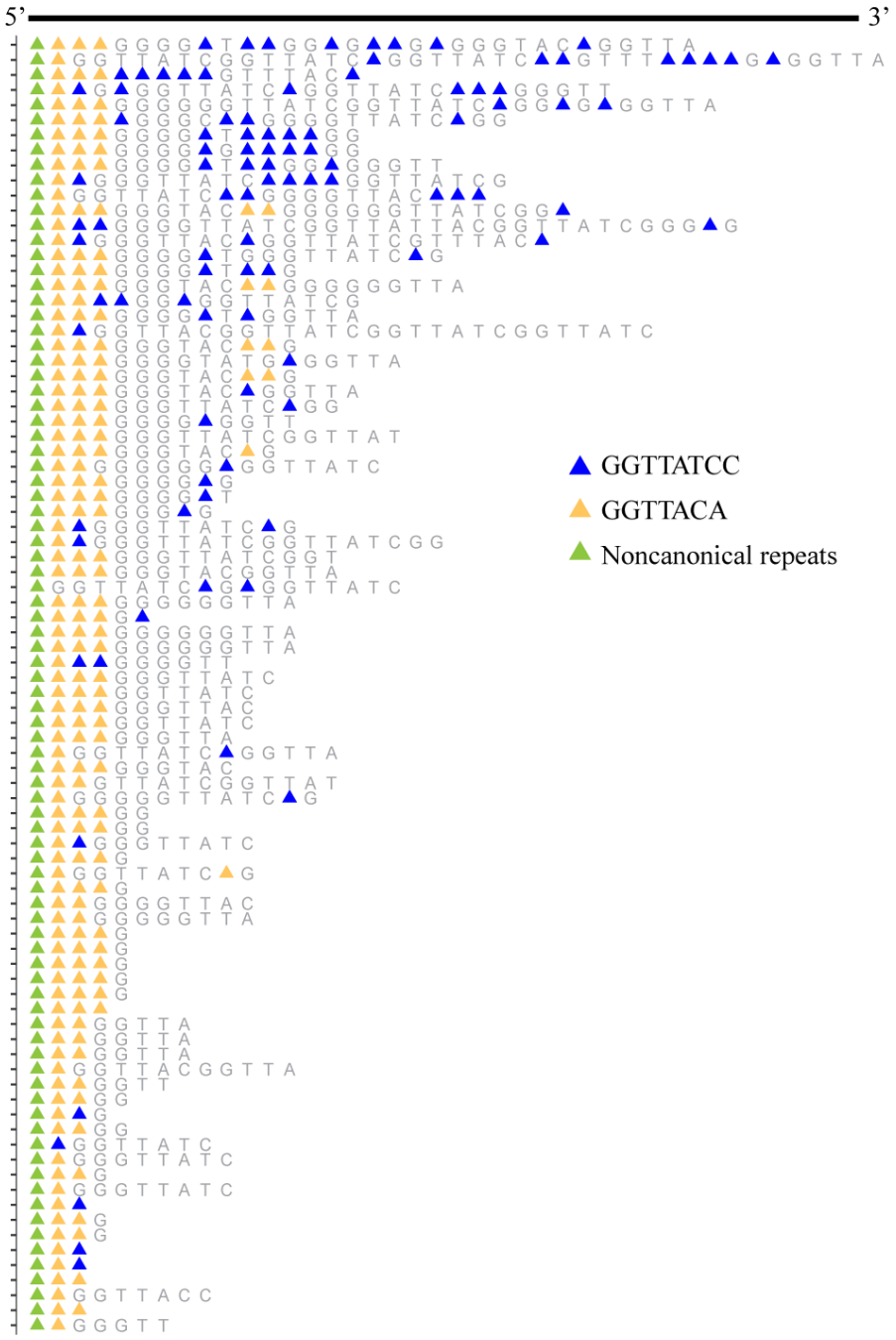


Figure 3.18 The GGTTATCC sequence is the most frequent motif in the 3'-CCAAU<u>AG</u>-5' template strain. Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-CCAAU**<u>AG</u>**-5'. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' repeats (gold triangle) are clustered toward the proximal end of the telomere. The most frequent motif GGTTATCC is highlighted by a blue triangle.

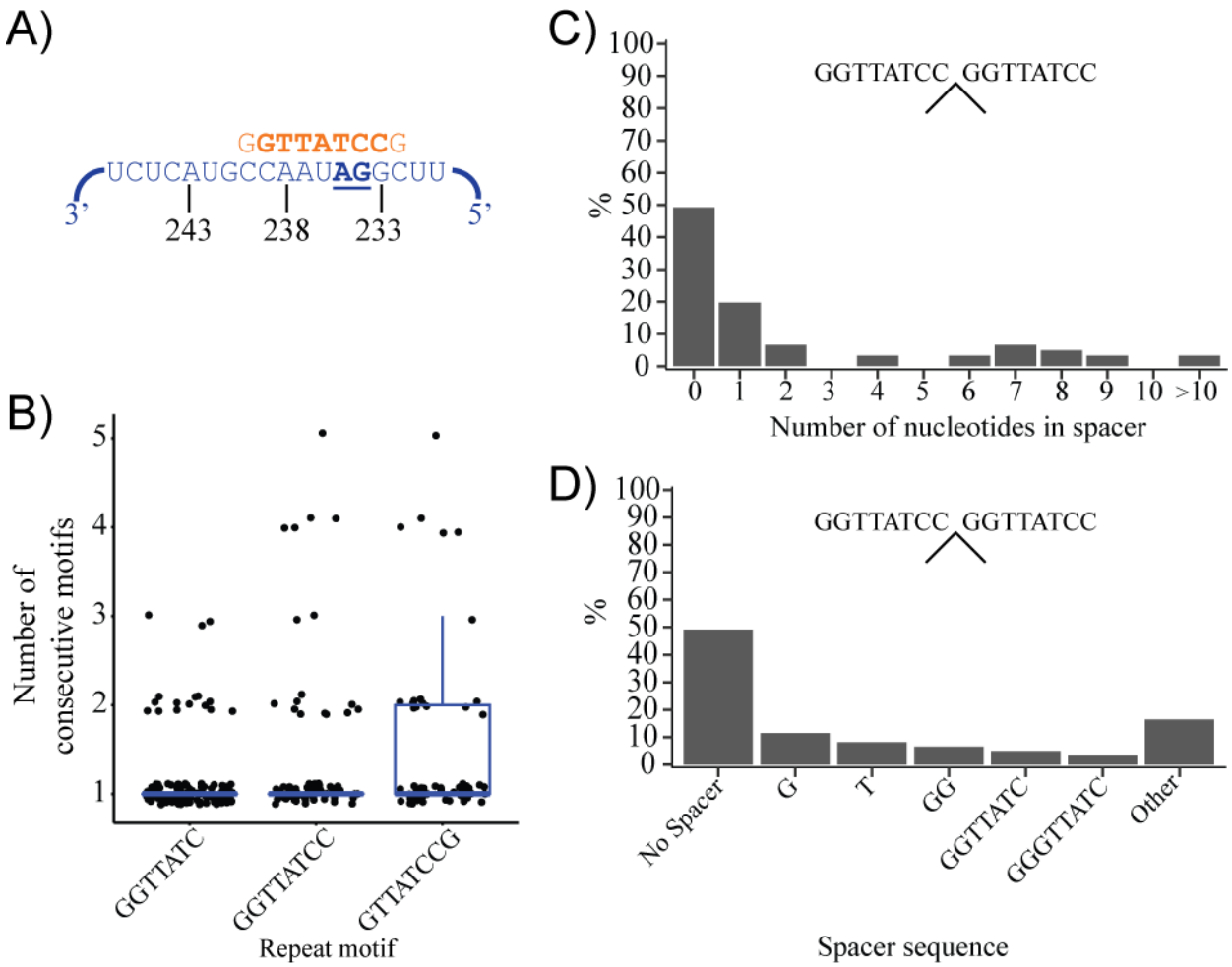


Figure 3.19 Permutations of GGTTATCC motif are interrupted by incomplete repeat sequences. Sequences of 86 telomeres (4,990nts combined). (A) Alignment of 3'-CCAAU**AG**-5' TER1 (blue, mutation bolded and underlined) to permutations (light orange) of the alternative repeat (orange). (B) Sequencing data were analyzed for the most frequent and repetitive motifs. Stretches of motifs were counted independently. Sequencing data were processed into GGTTATCC repeats and analyzed for the length of sequences between two repeats (C) and the sequence of nucleotides between two repeats; spacer sequences with frequency <2% are combined in the "other" category (D).

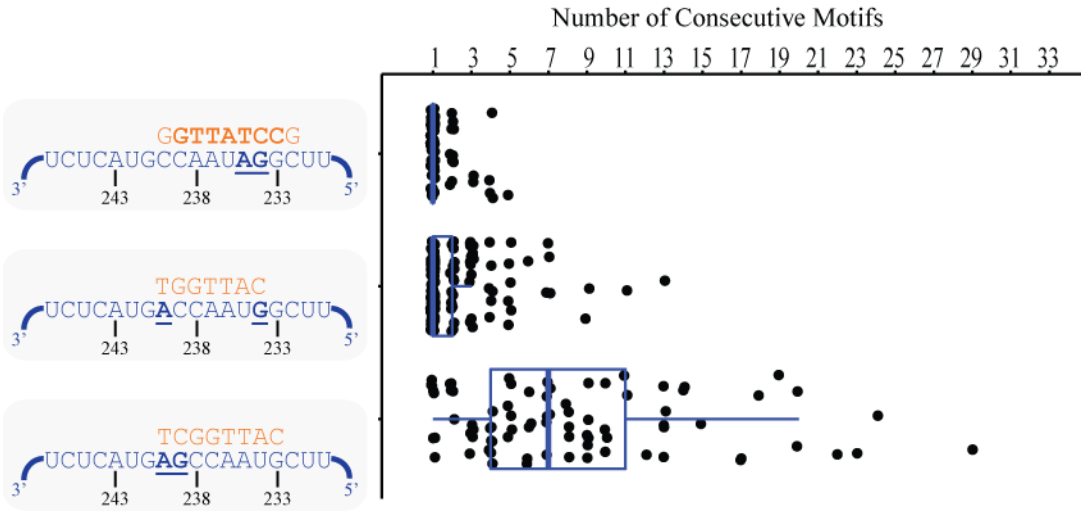


Figure 3.20 Example 2: Highly consecutive repeats end in 5'-TAC-3' templated by 3'-AUG-5' in the alternative TER1 template. The most frequent and repetitive motif (orange) of each strain was used to count the number of consecutive motifs within a stretch of repeats. Each stretch of repeats was counted independently in each telomere. TER1 (blue); variable nucleotides (bolded and underlined).

Example 3: Variable nucleotides 3'-GU-5'

Lastly, to determine if these same principles applied to the variable two nucleotides of the wild type template, we analyzed a third set of templates with 3'-GU-5' variable nucleotides. The repeat of the 3'-**G**CCAAU**U**-5'template strain was heterogeneous. The most frequent and repetitive templated repeats were permutations of 5'-GGTTACC-3', 5'-CGGTTAA-3', and 5'-GGTTAACC-3' (Figure 3.21A,B). Each of these motifs shared a 5'-CGGTTA-3' core. Unlike strains with more consistent repeats, analysis of nucleotides between core sequences revealed an increased percentage of spacers greater than 3nts in length (Figure 3.21C). The core motif was most often preceded by a single cytosine (~41%) or followed by a single adenine (~22%) (Figure 3.21D). Based on the spacer sequence analysis, the repeat of this strain is best described as 5'-$C_{0\text{-}1}CGG_{1\text{-}2}TT_{1\text{-}2}AA_{1\text{-}2}$. The template strain 3'-**GU**CCAAU-5' demonstrated a more stereotyped

repeat pattern defined by 5'-CAGGTTAC-3' repeats (Figure 3.22) The repeat was defined as 5'-CAGGTTAC-3' rather than 5'-GGTTACCA-3' because, although both yielded the same profile for consecutive repeat number and a maximum of eight repeats, 5'-CAGGTTAC-3' could best be explained by the actual template sequence (Figure 3.23A,B). This permutation overlaps with the wild type 5'-GGTTACA-3' sequence causing an increase in the number of single motifs and obscuring the analysis of spacer sequences between alternative TER1 templated repeats. The effect of introducing a 3'-AUG-5' into the template with wild type variable nucleotides resulted in a subtle increase in consecutive repeat number (Figure 3.24) compared to the more drastic changes found in the above examples 1 (Figure 3.13) and 2 (Figure 3.20). Importantly, although, wild type has a 3'-AUG-5' in the template it generates only 2-4 consecutive repeats due to its preference for the 5'-GGTTACA-3' sequence. These findings underscore the complexity of repeat addition in *S. pombe* and suggest the involvement of additional factors beyond the template.

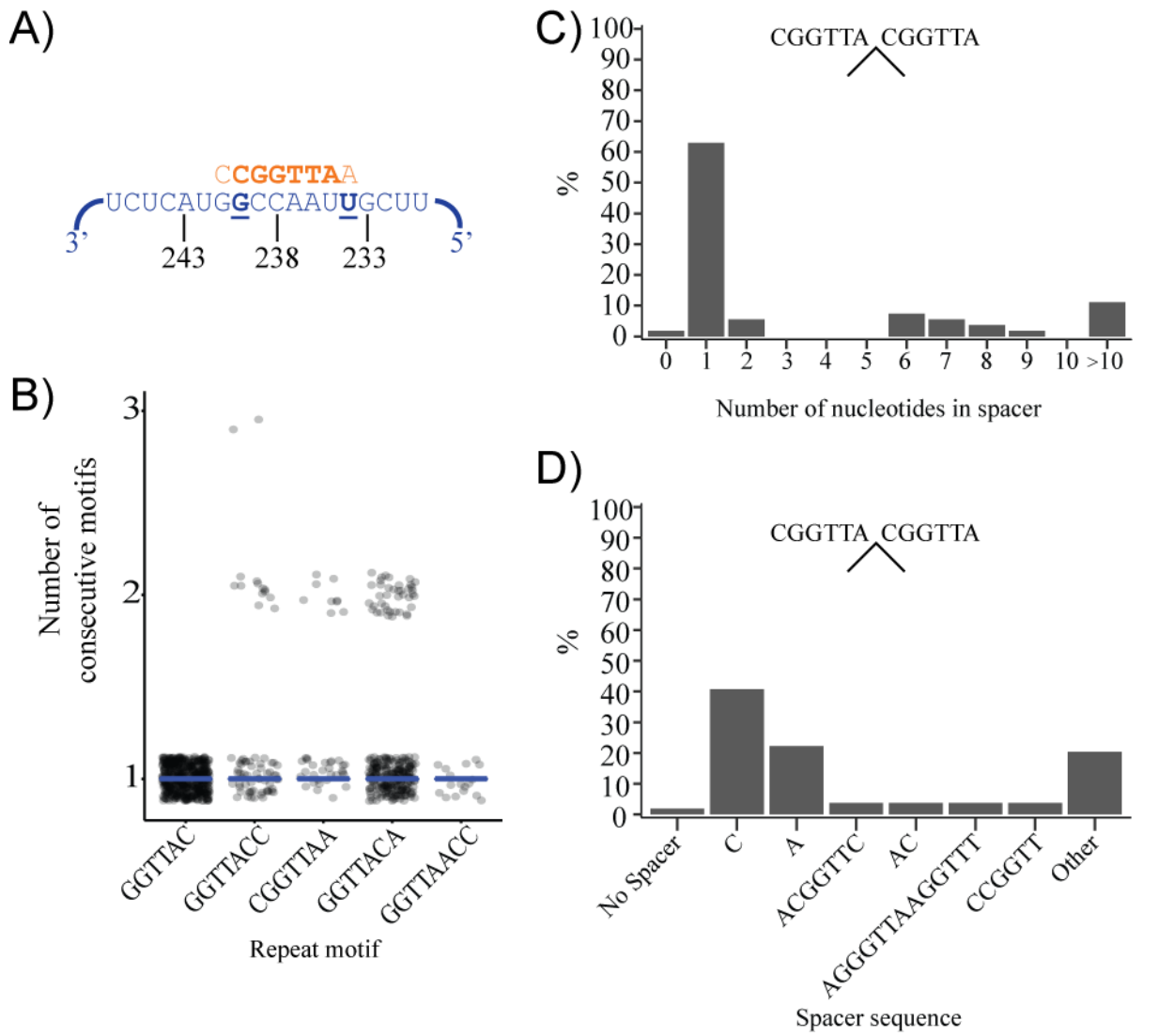


Figure 3.21 Heterogeneous telomeres of the 3'-GCCAAUU-5' template strain contain a 5'-CGGTTA-3' core repeat. Sequences of 90 telomeres (5,132nts combined). (A) Alignment of 3'-GCCAAUU-5' TER1 (blue; mutation bolded and underlined) to the newly synthesized core repeat (bold orange) and variations (light orange). (B) Sequencing data were analyzed for motifs that generated the highest number of consecutive repeats. Stretches of motifs were counted independently. Sequencing data were processed into CGGTTA repeats and analyzed for the length of sequences between two repeats (C) and the sequence of nucleotides between two repeats; spacer sequences with a frequency <2% were combined in the "other" category (D).

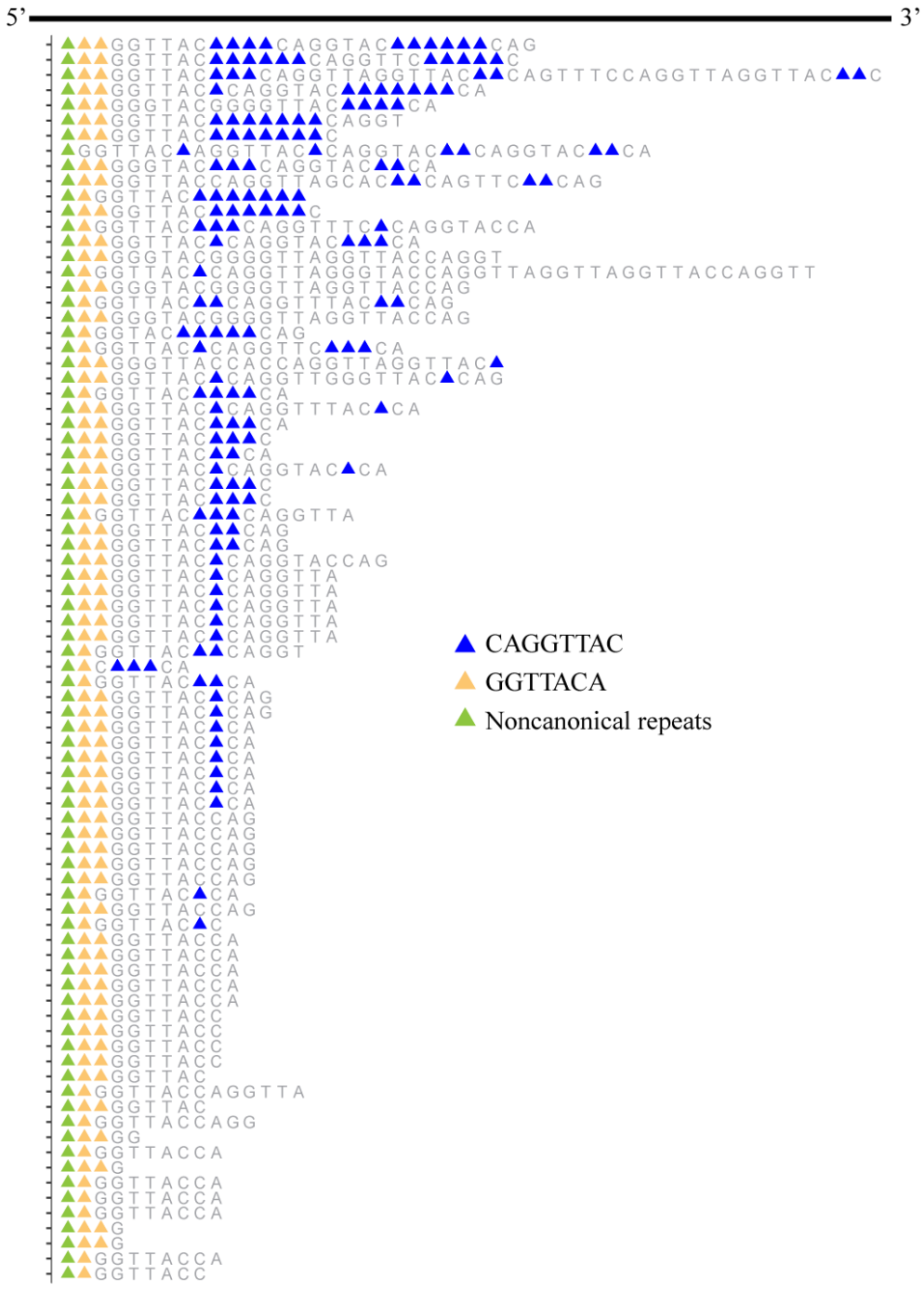


Figure 3.22 5'-CAGGTTAC-3' is the most common and consecutive repeat generated from the template 3'-GUCCAAU-5'. Alignment of all cloned telomeres from the plasmid strain with mutant template 3'-GUCCAAU-5'. Remnants of the wild type telomere including noncanonical repeats (green triangle) and 5'-GGTTACA-3' repeats (gold triangle) are clustered toward the proximal end of the telomere. The most frequent repeat 5'-CAGGTTAC-3' is highlighted by a blue triangle.

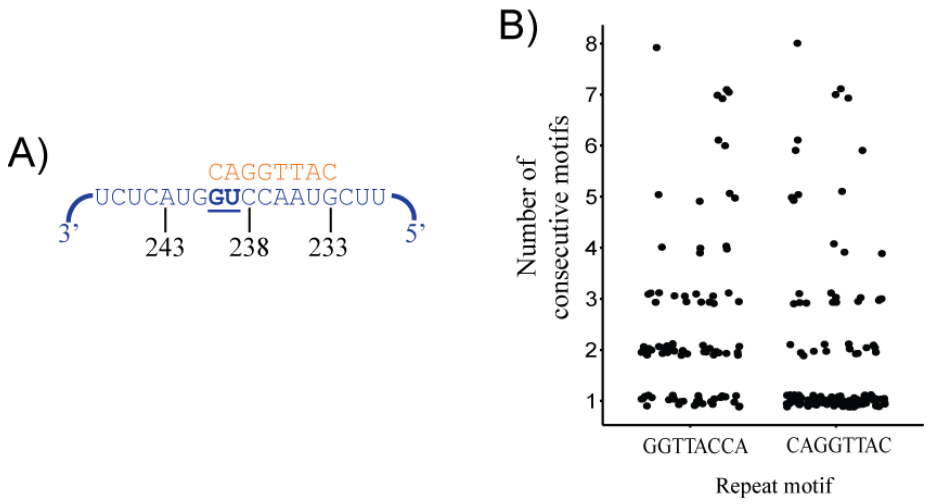


Figure 3.23 Alternative TER1 template corresponds to the 5'-CAGGTTAC-3' permutation of variant repeats. Sequences of 82 telomeres (5,277nts combined) (A) Alignment of 3'-**GU**CCAAU-5' TER1 (blue; mutation bolded and underlined) to the newly synthesized repeat (orange). (B) Sequencing data were analyzed for motifs that generated the highest number of consecutive repeats. Two permutations of the most frequent sequence are shown.

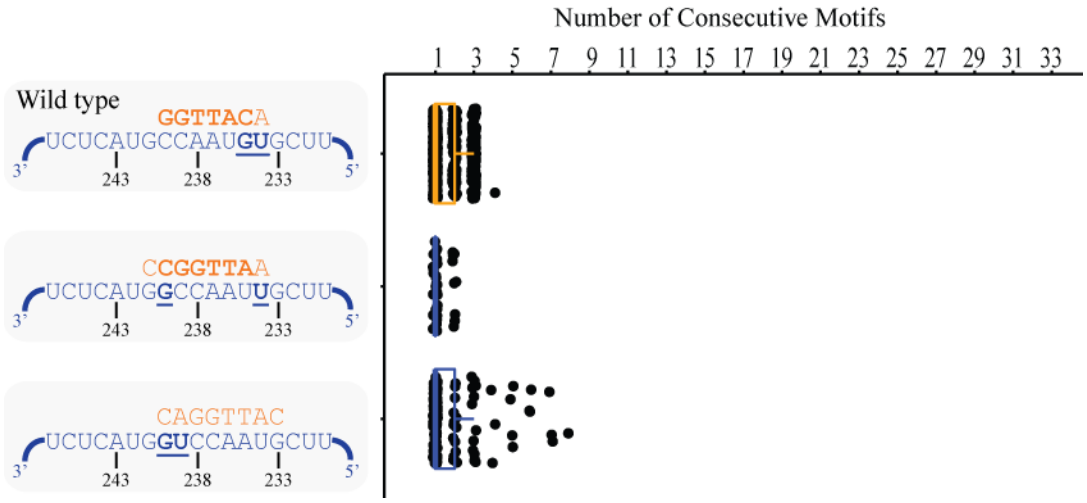


Figure 3.24 Example 3: Consecutive repeats end in 5'-TAC-3' templated by 3'-AUG-5' in the alternative TER1 template. The most frequent and repetitive motif (orange) of each strain was used to count the number of consecutive motifs within a stretch of repeats. Each stretch of repeats was counted independently in each telomere. TER1 (blue); variable nucleotides (bolded and underlined).

In conclusion, here we analyzed 3 examples of template permutations that conferred an increase in consecutive repeat number. With the exception of wild type, strains with 3'-AUG-5' in the template synthesized repeats ending in 5'-TAC-3' and strains without 3'-AUG-5' in the template did not. The latter tended to resemble wild type in the heterogeneity of repeats with a 5'-GGTTA-3' core.

III.4.6 Growth advantage of template is independent of telomere homogeneity

Through our analysis of nineteen alternative template strains and wild type, we identified six strains with increased consecutive repeat number compared to wild type 5'-GGTTACA-3' repeats (Figure 3.25A). While fission and budding yeast are known for heterogeneous telomeres, all vertebrate telomeres are direct 5'-GGTTAG-3' repeats. To determine whether homogeneous repeats resulted in a growth disadvantage in *S. pombe*, we compared the percent of reads for each template after 63 rounds of the 48 template in-flask competition experiment. Templates 3'-CCAAU**GC**-5', 3'-**AG**CCAAU-5', and CCAAU**AU**-5' outcompeted most of the other 5'-CCAAU-3' templates (Figure 3.25B). The other three templates, 3'-**GC**CCAAU-5', 3'-**A**CCAAU**G**-5', and 3'-**GU**CCAAU-5', were represented by less than 0.5% of reads in both cultures. Additionally, these less competitive templates demonstrated reduced telomere length after ~75 generations (Figure 3.25C). Templates, 3'-**A**CCAAU**G**-5' and 3'-**GU**CCAAU-5'also resulted in H1-H1' and H3-H3' end fusions consistent with the severe telomere shortening (Figure 3.25D). These results indicate that the competitive growth of *S. pombe* cells and their ability to maintain telomere function is irrespective of perfect repeat number.

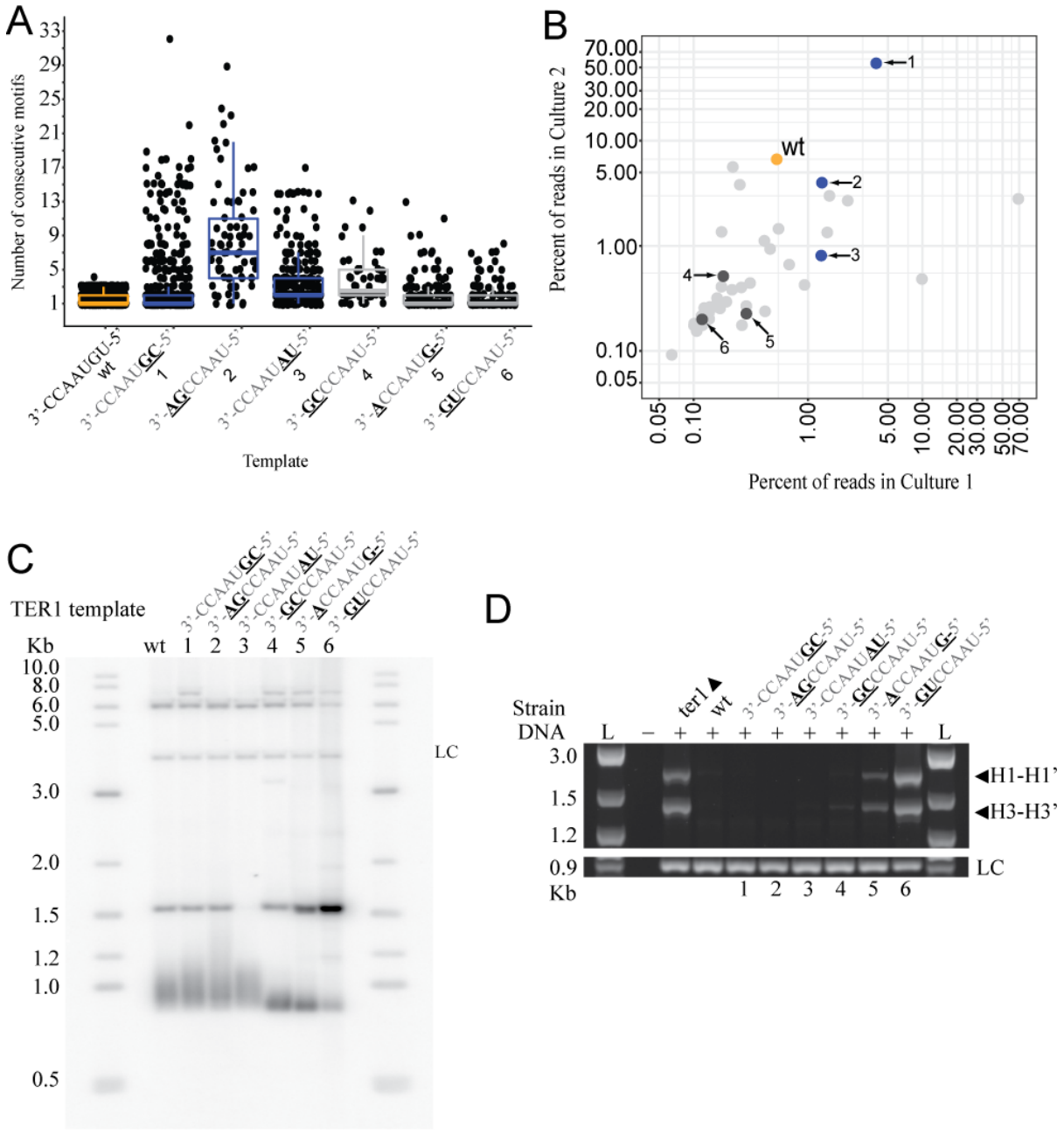


Figure 3.25 Competitive growth of template is independent of telomere homogeneity *in S. pombe*. Template numbers are used consistently throughout the figure. (A) Compiled data of 6 template strains with the highest consecutive repeat numbers compared to wild type (wt). (B) Frequency of each template after 63 rounds of competitive growth among 48 5'-CCAAU-3' TER1 template strains. Read counts were normalized to the total number of reads per library from Culture 1 or Culture 2. Numbers of each point correspond to the template numbers in A. (C) Telomere length analysis by southern blot hybridized with a probe specific for the subtelomeric element (STE). A probe specific to rad16 was used as the loading control (LC). (D) PCR to detect fused chromosome ends involving homology regions H1 and H1' and H3 and H3'. The catalytic subunit of telomerase *trt1* was used as a loading control (LC).

III.4.7 Repeat length varies among randomized seven nucleotide template strains to favor a 5'-TAC-3' telomeric ending

A repeat ending in 5'-TAC-3' is the common feature among strains with long stretches of successive repeats. This ending is templated by a 3'-AUG-5' in the respective alternative template TER1. Surprisingly, all but one of the templates correspond to repeat motifs that are either shorter or longer than expected from a seven nucleotide template. Template 3'-**A**CCAAU**G**-5' presents the exception as the theoretical seven nucleotide repeat was observed *in vivo* (Figure 3.16, Figure 3.17). Conversely, the 3'-CCAAU**GC**-5' template results in a six nucleotide repeat (Figure 3.26A) that is found in all telomeres *in vivo* (Figure 3.27C). The templates, 3'-**AG**CCAAU-5' and 3'-**GU**CCAAU-5', are preferentially reverse transcribed as an eight-nucleotide repeat motif (Figure 3.26B,C) found in all telomeres *in vivo* (Figure 3.27C). Similarly, 3'-CCAAU**AU**-5' is also used to generate an eight nucleotide repeat (Figure 3.27A) and was found in 99.8% of all telomeres (Figure 3.27C). Lastly, 3'-**GC**CCAAU-5', produced an eight-nucleotide repeat (Figure 3.27B) in 63.3% of telomeres while 2.5% of telomeres had only the theoretical seven-nucleotide sequence (Figure 3.27C). These unexpected findings suggest that reverse transcription of the template specifically stops at positions within TER1 that facilitate consecutive repeat addition.

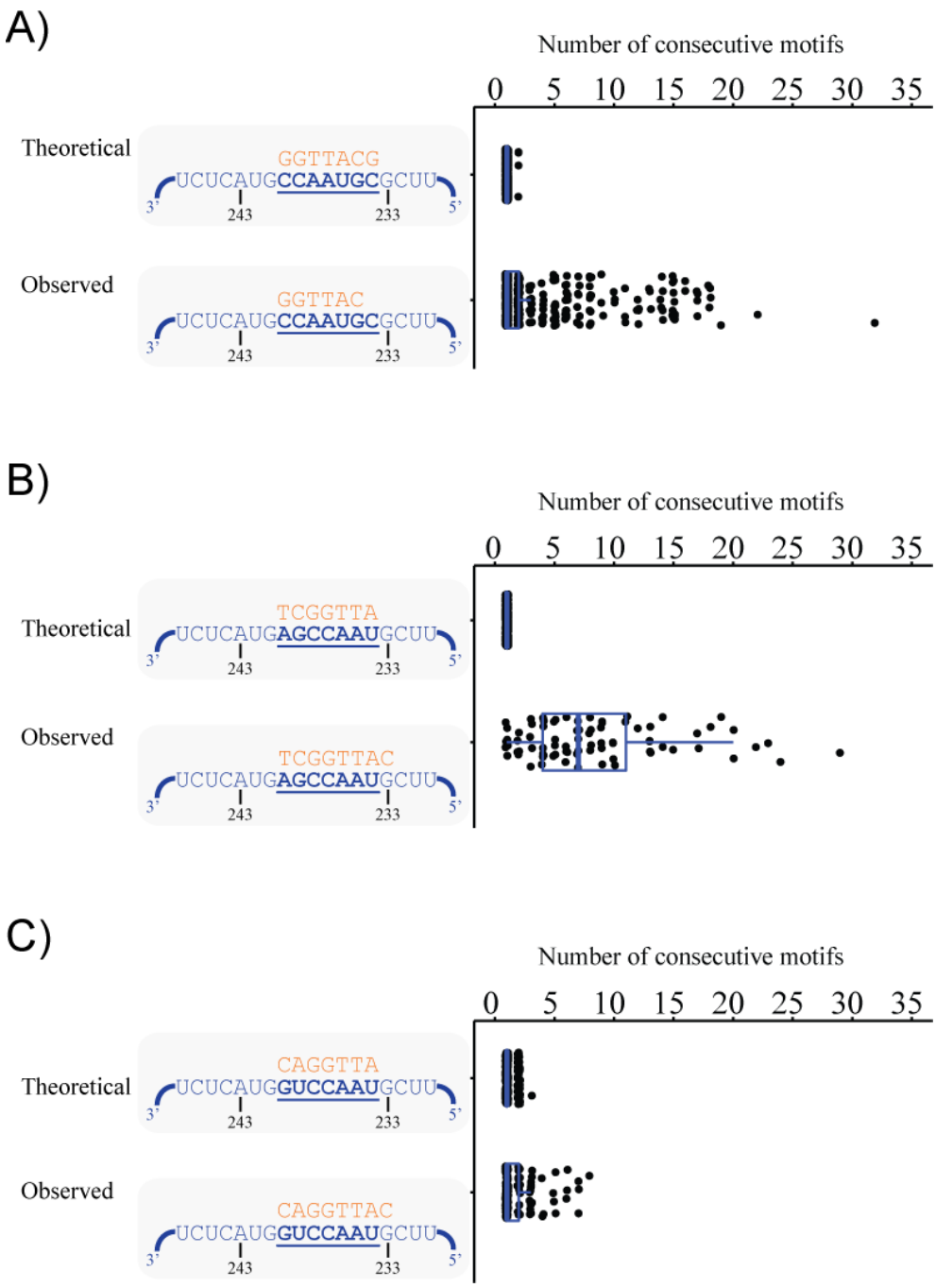


Figure 3.26 Observed repeat sequences *in vivo* differ in length from the expected repeat of a 7-nucleotide template. (A, B, C) Alignment of alternative template TER1 (blue; 7nt template bolded and underlined) to the templated repeat motif (orange). Theoretical sequences are based on a seven nucleotide template and *Mfold* boundary element predictions. Observed sequences were experimentally determined by cloning and sequencing telomeres using the G Overhang Capture Assay. Telomeres were analyzed for the number of consecutive occurrences of either the predicted or observed motif.

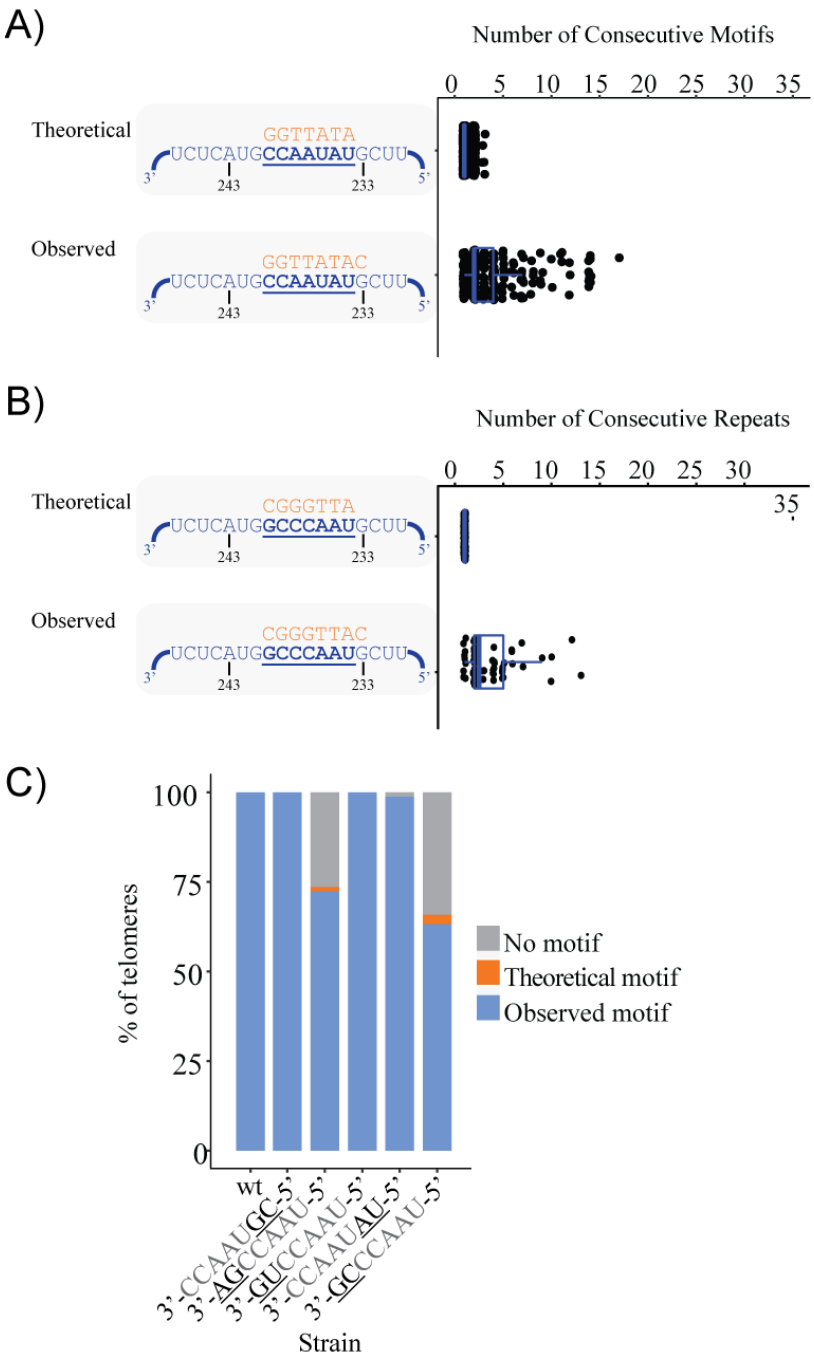


Figure 3.27 Predicted repeat motif is found in a minority of strains and telomeres. (A, B) Alignment of alternative template TER1 (blue; 7nt template bolded and underlined) to the templated repeat motif (orange). Theoretical sequences are based on a seven nucleotide template and *Mfold* boundary element predictions. Observed sequences were experimentally determined by cloning and sequencing telomeres using the G Overhang Capture Assay. Telomeres were analyzed for the number of consecutive occurrences of either the theoretical or observed motif. (C) Percent of all telomeres that had either the observed, theoretical, or neither motif. For wild type (WT) the GGTTACA motif was used.

III.4.8 Predicted structures of alternative template strains do not support a boundary element stop signal

Diverse telomerases use physical boundary elements or a pause signal within the telomerase RNA template to limit the nucleotide number of a single repeat (Chapter I.4). More specifically, wild type *S. pombe* TER1 forms a flexible hairpin structure at the 5' end of the template that most often results in a 7nt repeat [46]. To determine if a similar boundary element could contribute to the variable repeat lengths of the variant template strains, analysis was begun using *Mfold* software to predict RNA structures [228]. Nucleotide positions 1-1213 were entered into *Mfold* for unique alternative TER1 template sequences. To allow a direct comparison between wild type and the two template RNAs that generated the highest number of consecutive repeats (3'-CCAAU**GC**-5'and 3'-**AG**CCAAU-5'), minimal constraints were used to stipulate that 3'-CCAAU-5' of the template and the Sm/Lsm-binding site be single stranded. These stipulations regenerated published structures for the wild type RNA with free energies ~322 kcal/mol (Figure 3.28A) **[44-46]**. Regardless of template sequence, predicted structures for the variant templates included a boundary element that was inconsistent with the nucleotide addition of the observed repeat. These structures tended to maintain the global structure of wild type TER1 and also had free energies similar to the wild type closed conformation (Figure 3.28B, Figure 3.29A). However, a small number of predicted structures did support the nucleotide addition of the observed, perfect repeat. These structures tended to have a different global structure from wild type TER1 and free energies around -310 kcal/mol, generally increased from wild type (Figure 3.28C, Figure 3.29B). Importantly, despite the global change, the three-way-junction, or stem terminus element, which is bound by the catalytic subunit, is preserved. If the two arms of the wild type TER1 structure were constrained in addition to the minimum constraints, then boundary elements were again predicted for the variant templates. However,

they also precluded the addition of the last nucleotide of the observed repeat as G_{233} was base-paired in the boundary element (Figure 3.29C).

The interpretation of these findings is limited by the lack of input information for where the catalytic subunit and accessory proteins bind as well as complex, conserved folding structures such as the pseudoknot. Sill, in light of these preliminary findings, an attractive possibility is that the template confers a structural change to the RNA that results in the recruitment of a novel TER1 associated protein or alteration of the catalytic subunit-TER1 interaction to stabilize the RNA and adjust the catalytic cycle. Alternatively, the eight-nucleotide repeat strains may signify a departure from the wild type equilibrium that favors a closed conformation of the boundary element where a uracil is bulged (Figure 3.28A, closed boundary element). These eight-nucleotide repeats cannot be reverse transcribed in the closed conformation but may be reverse transcribed in the open conformation. However, the open conformation involves a G:U wobble that destabilizes the boundary element in wild type [46]. Additionally, a shift toward an open conformation cannot explain the six-nucleotide repeat of 3'-CCAAU**GC**-5'. Otherwise, a stop signal apart from a boundary element or other folding pattern must be considered.

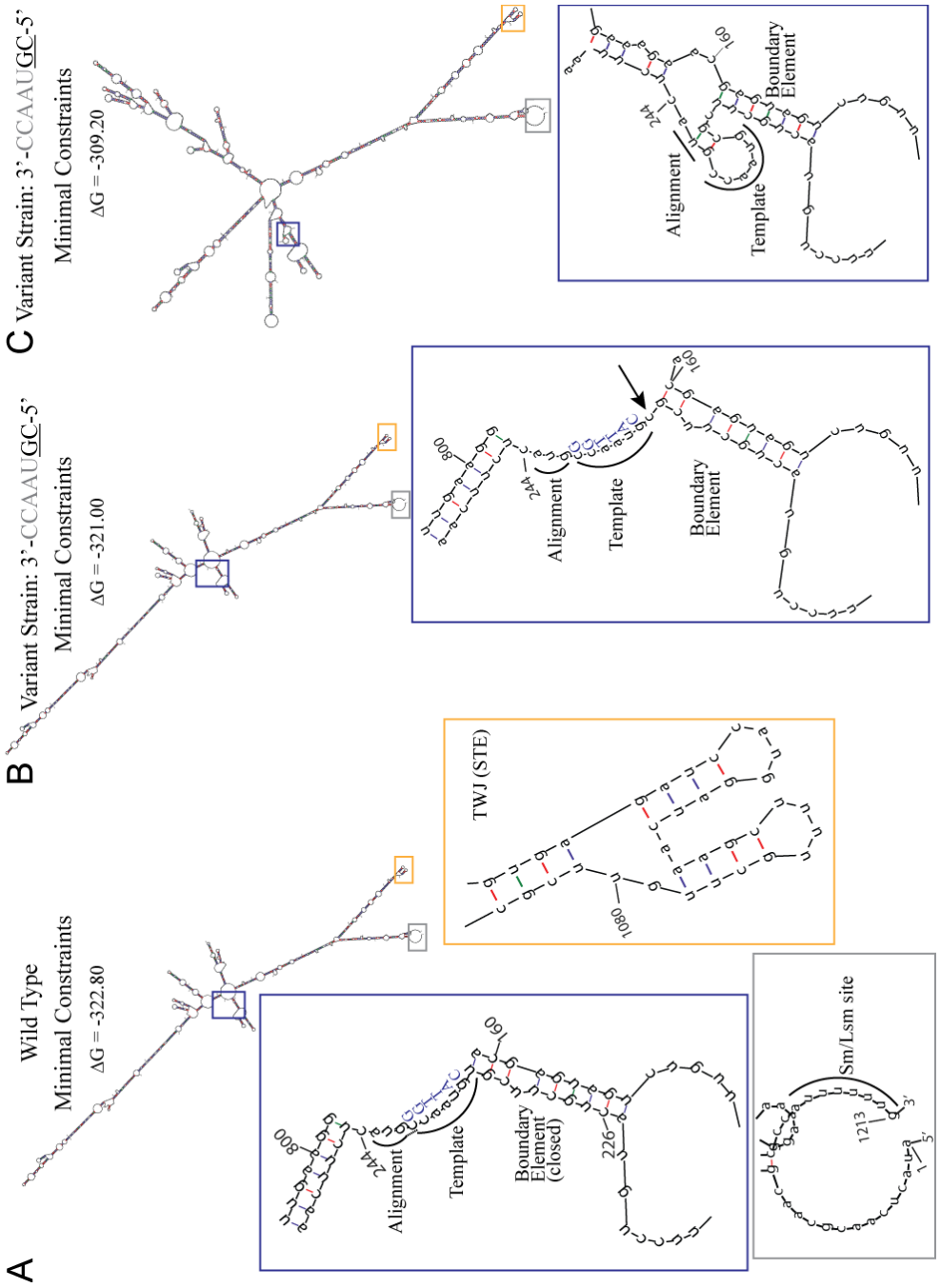


Figure 3.28 Predicted RNA structures of wild type and variant template 3'-CCAAUGC-5' TER1's. Minimal constrains include only requiring 3'-CCAAU-5' to be single stranded and the SM/Lsm binding site, 5'GUUUUUUA-3', to be single stranded. All ΔG values are in kcal/mol. A) 3 key structures of the wild type RNA are enlarged: 1) alignment, template with wild type repeat in Blue font, boundary element (BE) (Blue box), 2) Sm/Lsm binding site (Gray box), 3) Three-way junction (TWJ) or Stem Terminus Element (STE) (orange box). Same constrains were applied to 3'-CCAAU**GC**-5' TER1 (B, C). (B) Arrow denotes unpaired nucleotide that would potentially contribute to telomere repeat sequence.

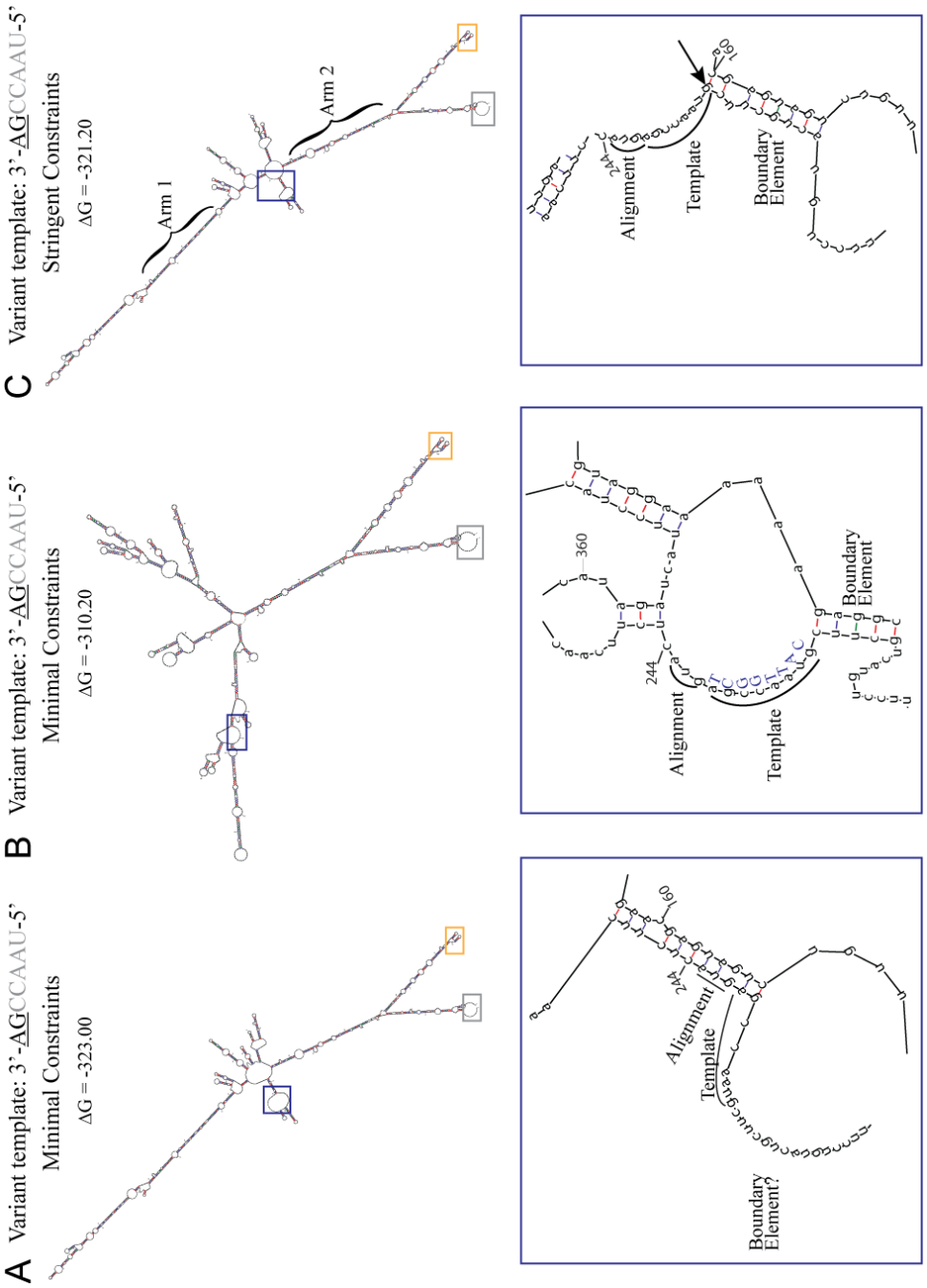


Figure 3.29 Predicted RNA structures of 3'-AGCCAAU-5' TER1. Same coloring scheme as in Figure 3.28. All ΔG values are in kcal/mol. Minimal constrains include only requiring 3'-CCAAU-5' to be single stranded and the SM/Lsm binding site, 5'GUUUUUUA-3', to be single stranded. (A, B). (B) Change in the global structure presents template that corresponds to the repeat sequence observed *in vivo* (Blue font). (C) Stringent constraints include stipulating that arms 1 and 2 of the wild type structure are base-paired in addition to the minimal constraints. Arrow points to single nucleotide use in repeat synthesis in vivo but base-paired in the boundary element of the predicted structure.

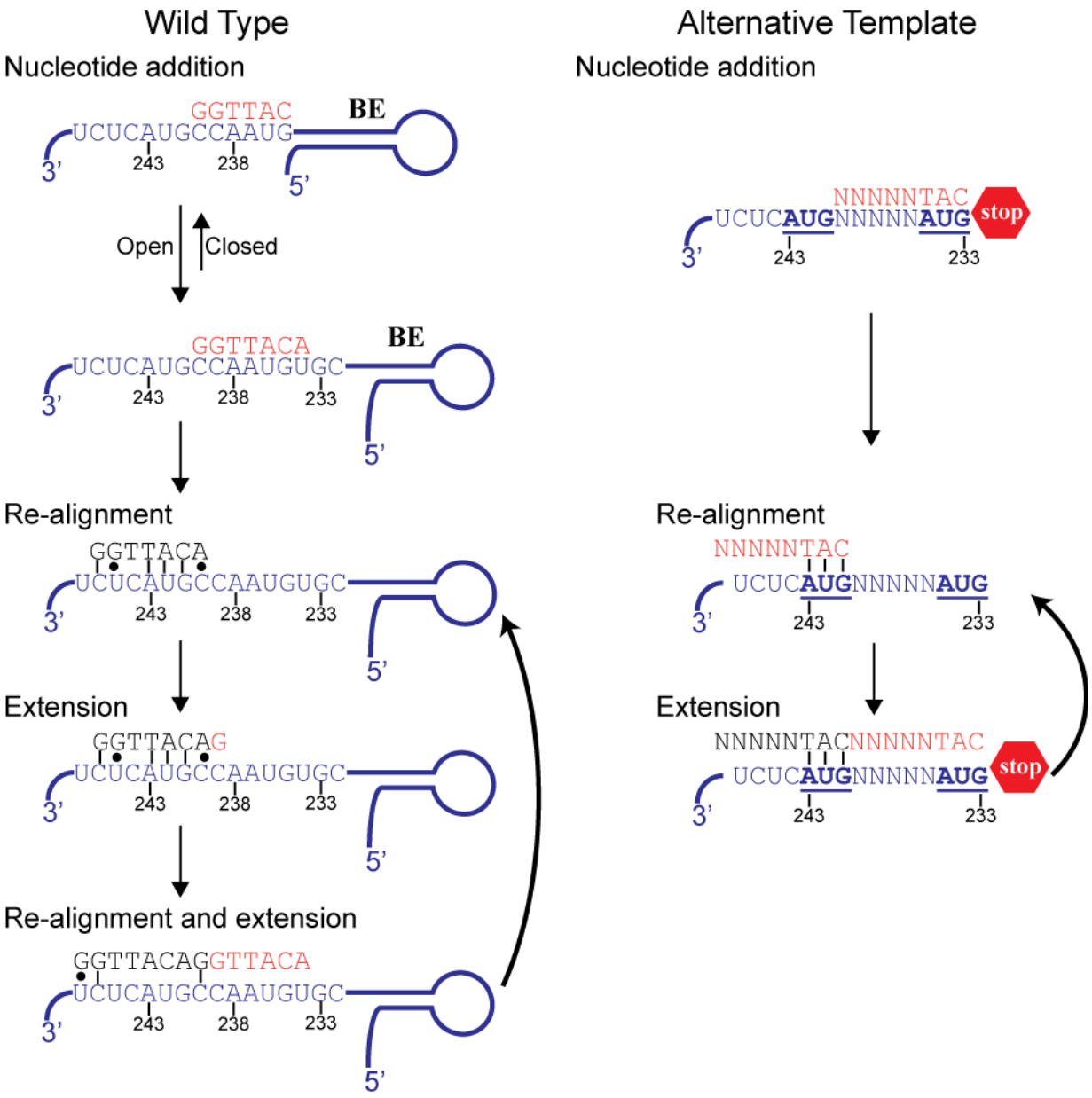


Figure 3.30 Model for telomere homogeneity in *S. pombe*. (Left) previously published model for repeat addition in wild type *S. pombe*. Process is shown beginning with nucleotide addition of repeat (orange) based on TER1 template sequence (blue) and either the closed or open conformation of the boundary element (BE). Two cycles of translocation and extension are depicted. Vertical lines denote canonical Watson Crick base pairing and a single dot signifies noncanonical base pairing. (Right) Proposed model for repeat addition in alternative template *S. pombe* with highly consecutive repeats. Random nucleotides are denoted with an "N". 3'-AUG-5' sequences in TER1 (blue, bold and underlined) together template a 5'-TAC-3' in the repeat (orange) that base pairs (vertical lines) with the alignment region of TER1. The process is shown beginning with nucleotide addition through extension and a second realignment.

III.5 Discussion

III.5.1 Summary

We have analyzed the sequences of functional telomeres from strains expressing different TER1 non-wild type templates. Surprisingly, this analysis revealed the usage of an alignment region shifted four nucleotides downstream of the wild type alignment region in the 3'-CCAAU**CU**-5' template strain. The 3'-CCAAU**CU**-5' template strain along with the 3'-UCCUCGG-5' template strain, described in Chapter II, provide growing evidence that *S. pombe* telomerase can switch alignment regions to facilitate templated nucleotide addition. We also observed a nearly total loss of telomere sequence heterogeneity among template strains that generated repeats ending in –TAC which was reverse transcribed from a 3'-AUG-5' in the variant template. These findings demonstrate that *S. pombe* telomerase can use a simpler catalytic cycle with a single translocation step and argue that it may behave as a processive enzyme, at least *in vivo*. Furthermore, we demonstrated that repeat homogeneity did not affect competitive growth in rich media and therefore does not present an apparent disadvantage to *S. pombe* under these conditions.

III.5.2 *S. pombe* shifted alignment indicates altered telomerase-telomere interactions

The present study of *S. pombe* non-wild type templates signifies altered interactions among telomeric DNA, telomerase RNA, and the telomerase holoenzyme as compared to wild type template telomerases. The wild type *S. pombe* alignment region and template may be surrounded by fixed structures at the 3' and 5' ends. At the 3'end, the alignment region could be restricted by the short hairpin downstream of the alignment/template (Figure 3.28A). This fixture would lend further support for why the wild type alignment region only shifts within a 1-3 nt range toward the 5' end [40, 43, 45]. Structural analysis of the interaction between the telomerase-RNA binding domain (TRBD) of the catalytic subunit and this base-paired region of

TER1, as was done in *T. thermophila*, is needed to confirm a fixation point at the 3' end of the alignment/template [136]. At the 5' end, the template is restricted by a hairpin loop upstream called the template-boundary element (TBE) [46]. The fixation of regions immediately upstream and downstream of the alignment and template likens *S. pombe* telomerase activity to the accordion model originally proposed for *Tetrahymena* telomerase [123].

Conversely, our data from two variant template strains indicate a greater range of dynamic interactions within the telomerase holoenzyme and between the enzyme and telomeric DNA. The template strains, 3'-UCCUCGG-5' (Chapter II) and 3'-**UC**UAACC-5' (Chapter III), make use of an alignment region shifted three and four nucleotides, respectively, downstream of the wild type alignment region. Whereas wild type preferred 5' translocation sites, these strains frequently translocated toward the 3' side of the template to accommodate the 3'end of the variant-templated repeat. Unexpectedly, the 3'-**UC**UAACC-5' telomerase switched between the variant alignment and the wild type alignment, resulting in spacer sequences interspersed between 5'-GTACGGTTAGA-3' repeats (Figure 3.2). Because the sequence downstream of the template was not mutated, if the short hairpin at the 3' end of the alignment/template precluded the use of this sequence for alignment, then these telomerases would not have aligned the telomeric repeat to this position. Therefore, the use of a shifted alignment argues against a fixed structure necessitated in the accordion model and instead marks considerable plasticity in the telomerase holoenzyme-telomeric DNA interaction.

III.5.3 TER1 template and repeat addition processivity

Telomerase RNAs mediate catalytic activity in addition to providing a template for telomeric repeats. Direct interactions between the catalytic subunit and telomerase RNA have

been shown to promote repeat addition processivity (RAP) in T*etrahymena* [132] and human [129]. Mutation of the template in human telomerase impaired repeat extension and repeat translocation rates [229]. Similarly, mutation of the *Tetrahymena* alignment region or template decreased telomerase repeat addition *in vitro* [133]. We have demonstrated that mutation of the *S. pombe* template results in the incorporation of long stretches of consecutive, variant repeats rather than degenerative, wild type repeats. Similarly, although wild type *S. cerevisiae* telomerase is nonprocessive, replacement of the endogenous template with the human sequence resulted in stretches of human repeats ranging from 7 to 45 consecutive motifs [125, 230]. However, our findings additionally provide evidence that even subtle changes in the template sequence can influence the number of successive repeats. Notably, as minimal as a single nucleotide change could switch the repeat phenotype from degenerate to nearly perfect (Figure 3.7 compared to Figure 3.8). Furthermore, a change in the permutation of the template without a change in nucleotide composition also affected the range of direct repeats (Example 1 and Example 3 of section III.4.5). Our findings provide further evidence that the template of an otherwise nonprocessive enzyme can force direct repeats *in vivo,* indicating the possibility for processive activity by *S. pombe* telomerase.

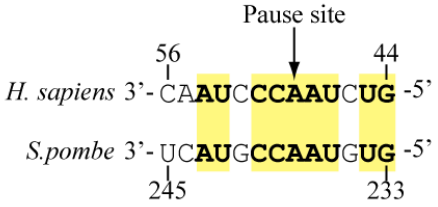
Increasing the affinity between telomerase RNA and telomeric DNA has also been shown to enhance translocation efficiency and facilitate RAP while decreasing the affinity impaired RAP [231-233]. We found that strains with the most consecutive repeats all harbored a 3' telomeric end that directly base paired with the wild type 3'-AUG-5' alignment sequence. Therefore, we favor a model for direct repeat synthesis in *S. pombe* where strong base pairing between the last 3' nucleotides of the telomeric repeat and the alignment region facilitate RAP. Additional experiments are needed to test whether *S. pombe* telomerase becomes processive in

the presence of variant templates. An alternative possibility is that the catalytic subunit is autonomously nonprocessive but becomes processive in the presence of additional *in vivo* factors. In T*etrahymena*, the wild type enzyme was processive *in vitro* and nonprocessive *in vivo* when mutant template telomerase RNAs were introduced [58, 234]. Additionally, the shelterin proteins, Pot1 and TPP1, also regulate telomerase processivity [128].

III.5.4 Does *S. pombe* share a conserved template pause site with human telomerase RNA?

The discovery of nearly perfect repeats in variant TER1 template strains revealed variation in the length of a single repeat motif. Theoretically, if the entire template was used to synthesize a repeat unit, a seven nucleotide sequence should have been observed. Yet, actual repeat units ranged from 6-8 nucleotides. One possible explanation is that the physical boundary element is adjusted from its wild type conformation to either limit or extend reverse transcription. However, *Mfold* software predicts that only structures that change the global folding pattern of the RNA support perfect repeat addition. Another possibility is that the alternative templates have unmasked a conserved pause site with human telomerase RNA (hTR). *Brown et al.* published data showing that the A_{49} of the hTR template represented a pause site such that only three nucleotides were added following its reverse transcription [127]. Intriguingly, five of the six direct repeat-variant template strains

A



B

Strain	TER1	Repeat motif length
FP1431	3'- UCAUGCCAAUGCGCUU -5'	6nt
FP1433	3'- UCAUGACCAAUGGCUU -5'	7nt
FP1421	3'- UCAUGAGCCAAUGCUU -5'	8nt
FP1464	3'- UCAUGGCCCAAUGCUU -5'	8nt
FP1463	3'- UCAUGGUCCAAUGCUU -5'	8nt

Figure 3.31 Potential pause site in variant TER1 templates. (A) Comparison of telomerase RNA sequences. (B) The five direct repeat-variant template strains with repeat lengths that can be explained by a potential pause site. The template that is used to reverse transcribe the repeat motif is underlined. Potential pause site is bolded.

share sequence homology with this region of hTR and exhibit repeat lengths that can be explained by a putative pause site (Figure 3.31). Although the presence of a pause site within these *S. pombe* strains needs to be confirmed, its existence would provide an exciting opportunity to study the properties of a pause site in a more genetically tractable model system.

A genetics approach can be employed to test whether a boundary element or pause site is contributing to the variation in lengths of single repeat units. The length of a repeat motif would change predictably if point mutations that disrupted the presumptive boundary element, or base pairing within the hairpin structure, were introduced. Then, compensatory mutations on the complementary strand that rescue the originally observed length would confirm a physical boundary element upstream of the template. Conversely, if the presumptive pause site is mutated, then an increase in the length of repeat motifs would suggest disruption of the pause signal.

III.5.5 Model for telomere homogeneity in *S. pombe*

In wild type, open conformation of the boundary element predominates to generate mostly 5'-GGTTACA-3' repeats [40]. The 5'-ACA-3' ending results in two translocation steps to add the second repeat or multiple translocation steps to incorporate four to six guanines between repeats [46]. Similar to human telomerase RNA, alternative template *S. pombe* strains with homogeneous telomeres have a template sequence that matches the alignment region. Repeats therefore end in 5'-TAC-3'which directly base pairs with the 3'-AUG-5' in the alignment region, positioning the newly synthesized repeat in a +1 register that allows the second repeat to be reverse transcribed without multiple translocation steps (Figure 3.30, right). A stop signal between the catalytic subunit and the template or a processing event fixes the telomeric repeat length and ensures a 5'-TAC-3' ending.

This model explains how alterations to the telomerase RNA template can switch synthesis from degenerative repeats to consecutive repeats in *S. pombe*. Whereas wild type *S. pombe* telomerase may function in a similar way to the accordion model in *Tetrahymena*, our analysis of variant template *S. pombe* telomerases provides evidence in support of a distinct catalytic mechanism, resembling human telomerase. Importantly, this implicates *S. pombe* as an evolutionary transition between ciliate telomerase catalytic cycles and vertebrate telomerase catalytic cycles.

Chapter IV: Identification of templates causing rapid cellular senescence

IV.1 Abstract

Telomerase is a ribonucleoprotein complex that reverse transcribes a template region within its RNA subunit to extend chromosome ends. Telomere elongation prevents the gradual loss of DNA inherent to each replicative cycle and provides a scaffold onto which protective and regulatory proteins assemble. In the absence of telomerase, critically short telomeres become uncapped, rendering them vulnerable to DNA damage response signaling, check point activation, cellular senescence, and apoptosis. The telomerase catalytic subunit is overexpressed in at least 85% of all cancers (90% of breast cancers), allowing cancer cells to evade replicative senescence and proliferate. Thus, telomerase may be a viable target for anti-tumor therapy. Telomere destabilization exploits aberrant telomerase activity to introduce mutant repeats specifically into the telomeres of cancer cells. However, drug discovery has been hampered by the lack of information regarding the induction of a potent telomere uncapping response. To identify multiple telomerase RNA templates that can be used to mutate telomere sequences, we performed a competitive growth screen of all 16,384 possibilities of a randomized template in the fission yeast telomere model system. We report 481 templates that were consistently lost after an average of 18.1 generations in triplicate experiments. A consensus sequence of alternating GC nucleotides was found among these templates. Our findings provide a foundation for the optimization of telomerase RNA templates that can be used to mutate and destabilize telomeres in telomerase positive cancer cells.

IV.2 Introduction

Telomere destabilization is one approach to anti-cancer therapy. Destabilization could be achieved through telomere uncapping where the telomerase RNA (hTR) template is manipulated to produce sequences at the chromosome end that result in the removal of the sequence-specific, telomere-associated, protective proteins. Studies have demonstrated that a $p53^{+/+}$ background is required for telomere uncapping to confer cell death in epithelial cancers like breast carcinomas [147, 235]. Telomere uncapping elicited by point mutations in the hTR template has been shown to effectively elicit cell death in human cancer cell lines [236-238]. Whereas, these studies collectively focused on two templates that were predicted to generate mutant repeats, the present study aimed to survey all template possibilities for those that generated the most immediate growth defect. Therefore, the experiments described in this chapter use a more comprehensive approach of randomizing the entire telomerase RNA template. In a competitive growth experiment similar to that described in Chapter II, cells that are unable to maintain telomere homeostasis will enter senescence and the templates contained in them will become the least abundant, or losers of this competition. By studying the earliest template losers, we have identified mutant templates whose telomere phenotypes can now be characterized and have uncovered shared nucleotides patterns that are strongly selected against.

IV.3 Materials and Methods

IV.3.1 *S. pombe* Library Generation and Competition Time Course

The diploid strain PP407 [46] was sporulated on Malt Extract agar at 28°C for 48 hours and subjected to glusulase (Perkin Elmer) treatment then *ter1*$^{-}$ haploid cells were selected on YEA 100μg/ml Geneticin plates. Patches from three separate *ter1*$^{-}$, *ade-M216* haploid colonies were used to propagate separate 100ml YES cultures into log phase (0.5 to 1.0 x 10^7 cells/ml). Cultures were then washed twice with chilled 1.2M sorbitol (Sigma) and resuspended at 1 x 10^9

cells/ml. Cells (2 x 10^8) were electroporated with 1μg of pMP04 plasmid library in a chilled 0.2cm electrode gap cuvette (VWR) using a BioRad Gene Pulser II set to 2.25kV, 200 Ohm, 25μF. Transformation reaction volumes were increased to 1ml with 1.2M sorbitol, centrifuged in a barrel centrifuge for 5 minutes, and resuspended in 1.25ml YES. Three sets of four transformations were pooled for a total of three 5ml cultures (Pool 1, 2, 3) and recovered at room temperature in the shaker at 250 revolutions per minute (rpm) for 6 hours. The volume of each recovery culture was gradually increased over 23 hours to 100ml YES plus ClonNat (50μg/ml). Then, 100ml cultures were grown to stationary phase for an additional 24 hours. The stationary phase culture was collected as 'Round 1' and used to dilute 'Round 2' culture to 5 x 10^4 cells/ml. All subsequent cultures were grown to log phase and diluted to 5 x 10^4 cells/ml to propagate the subsequent round. Final cultures were divided into aliquots for glycerol stocks for odd rounds only and genomic DNA extractions for all rounds. To determine library coverage, transformation efficiency (TE) for each pool was calculated (Equation 5) from plating the equivalent of 1.78μl from the 4 combined transformations and 4μg of DNA. Approximately 800-1100 transformants were counted from triplicate plates for each pool and averaged.

Equation 5: $$TE = \frac{\text{Number of transformants x Volume of cells used for transformation}}{\text{Volume of cells plated x Amount of DNA } (\mu g) \text{for transformation}}$$

IV.3.2 Illumina library Preparation, Sequencing, and Computational Analysis

Libraries were generated for the odd rounds of each pool by PCR with custom, Illumina-compatible primers as described previously for the pMP76- pMP123 rich media competition (Chapter II). All custom PCR libraries were combined into a 10 nM pool and spiked into complex genomic libraries at 15%. Samples were run in Rapid mode of the Illumina HiSeq 2500 to yield 50 bp single reads. Raw data were processed as described in Chapter II without a

minimum read filter. Calculations of fitness, total number of wild type generations, and total number of population generations were done according to the formulas described in Chapter II.

IV.4 Results

IV.4.1 Modified competition experiment to identify template losers

A shorter competition time course was employed to screen for template losers. In this competition, cultures were grown to log phase (0.9-1.4 x 10^7 cells/ml) prior to dilution to more robustly select against template mutants that cause a defect in telomere maintenance (Figure 4.1A). The presence or absence of templates was monitored after Round 1, 3, and 5 of three independent cultures. Filtering was conducted as described in Chapter II with the exception of a minimum reads filter. This filter was omitted to allow for the identification of low frequency templates. The number of reads per library decreased as the complexity of the PCR library also decreased. However, after the final filter step of Round 5 libraries, read counts were between 6.0 and 11.8x the complexity of the initial library (Figure 4.1B). The number of wild type generations was determined from the percent of wild type template reads and culture densities after each round (Chapter II). The total number of population doublings was determined directly from counting culture densities at stationary phase after each round. By the final time point, wild type had grown 27.4, 28.0, and 28.2 generations in cultures 1, 2 and 3, respectively (Figure 4.1C). Growth of the entire population consistently lagged behind wild type ending at 24.9, 24.1, and 25.6 generations in cultures 1, 2 and 3, respectively. We next characterized the total number of templates after each round. The *ter1* template plasmid library was previously determined to have all 16,384 possibilities of a randomized 7nt sequence (Chapter II). After the initial time point, 16,382 to 16,384 of these sequences were detected in each of the cultures (Figure 4.1D). The missing template sequences were absent from only one of the three cultures.

We observed a constant decrease in template diversity with each successive round consistent with a loss of cells with dysfunctional TER1 templates.

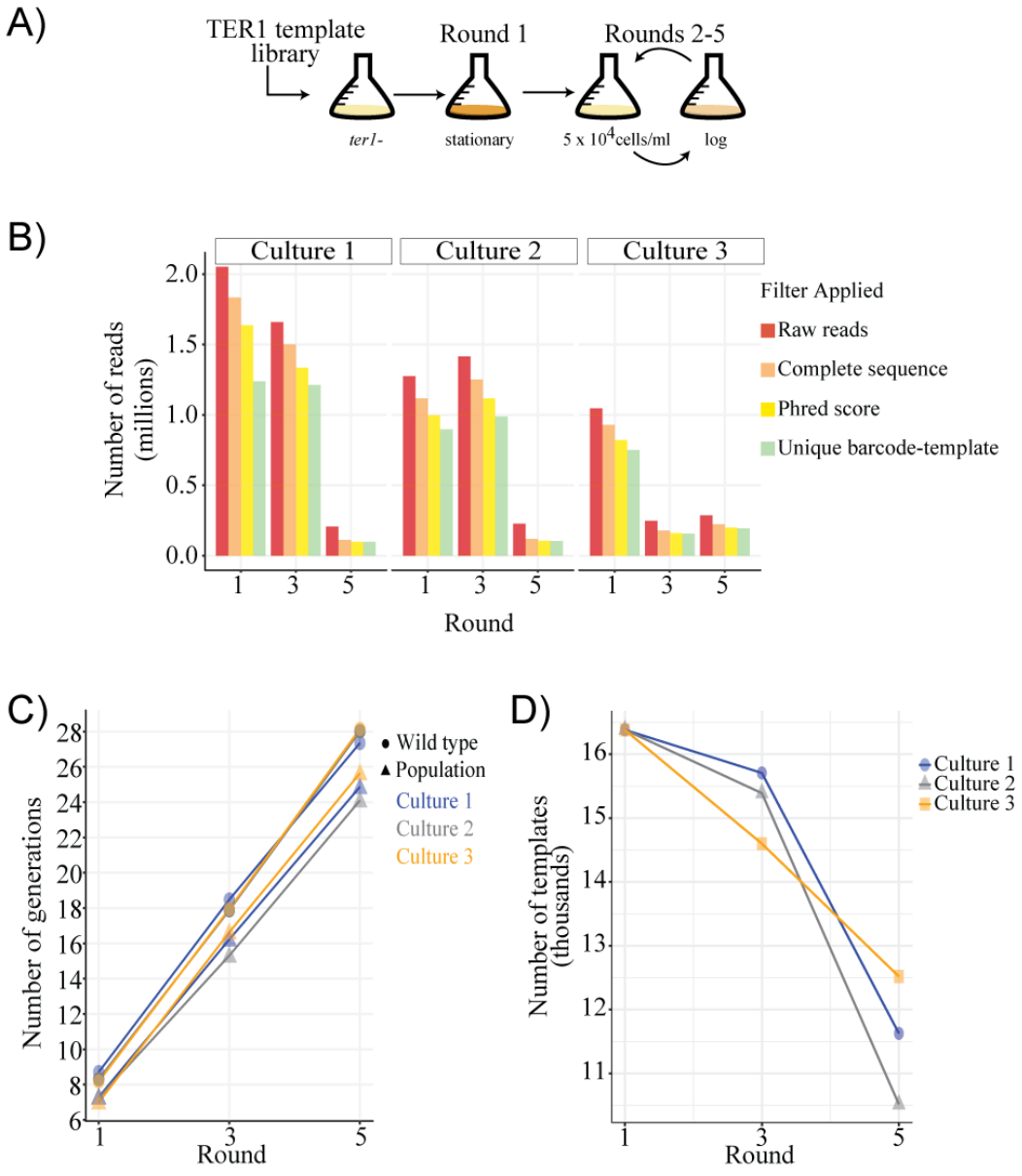


Figure 4.1 Decreased template diversity in a log-phase, short time course competition. (A) Experimental design. The pMP04 template library was transformed into haploid *ter1*$^{-}$ cells from PP407. Round 1 stationary phase cultures (6.4 -7.9 x10^7 cells/ml). Round 2-5 log phase cultures (9.0 – 1.4 x 10^7 cells/ml). (B) Number of reads after each filter step. Filter steps are organized on the x-axis in the order in which they were performed. (C) Total number of wild type generations were calculated based on the normalized number of wild type reads and culture densities. Total population number was calculated from culture densities. (D) Number of unique templates after each round.

IV.4.2 Sequence pattern among template losers

We next began to characterize the sequence of mutant templates. Template losers were defined as those sequences that were present at Round 1 but absent from Round 3 and Round 5. Because these templates had read counts at Round 1 but zero reads in Round 3 and Round 5, their fitness was zero after only 17.9 – 18.5 wild type generations (Figure 4.1C). The three cultures shared 481 templates that were lost after Round 1 (Figure 4.2A). The 5'-TAACC-3' containing rich media *ter1* template winners and the 5'-GGCTCCT-3' and 5'-TGAACCT-3' minimal media template winners were not present among the 481 templates. Within these sequences, an alternating CG pattern emerged (Figure 4.2B). The cytosine nucleotide predominated at positions 236, 238, and 240 of the template in the *ter1* gene accounting for 55.5%, 47.6%, and 51.6% of nucleotides at each position, respectively (Figure 4.2C). The guanine nucleotide demonstrated a more modest bias representing 35.1% and 39.3% of nucleotides at positions 237 and 239, respectively. Additionally, the vast majority of these 481 template mutants were found to have a fitness approximating zero in the minimal media experiments (Figure 4.3A, B). Only 189 and 264 of the 481 templates were present at Round 1 of the Culture 1 and Culture 2 minimal media experiments, respectively. The remaining templates were below the detection limit. These findings suggest that either a C or G nucleotide is selected against at specific positions within the *ter1* template in both rich and minimal media.

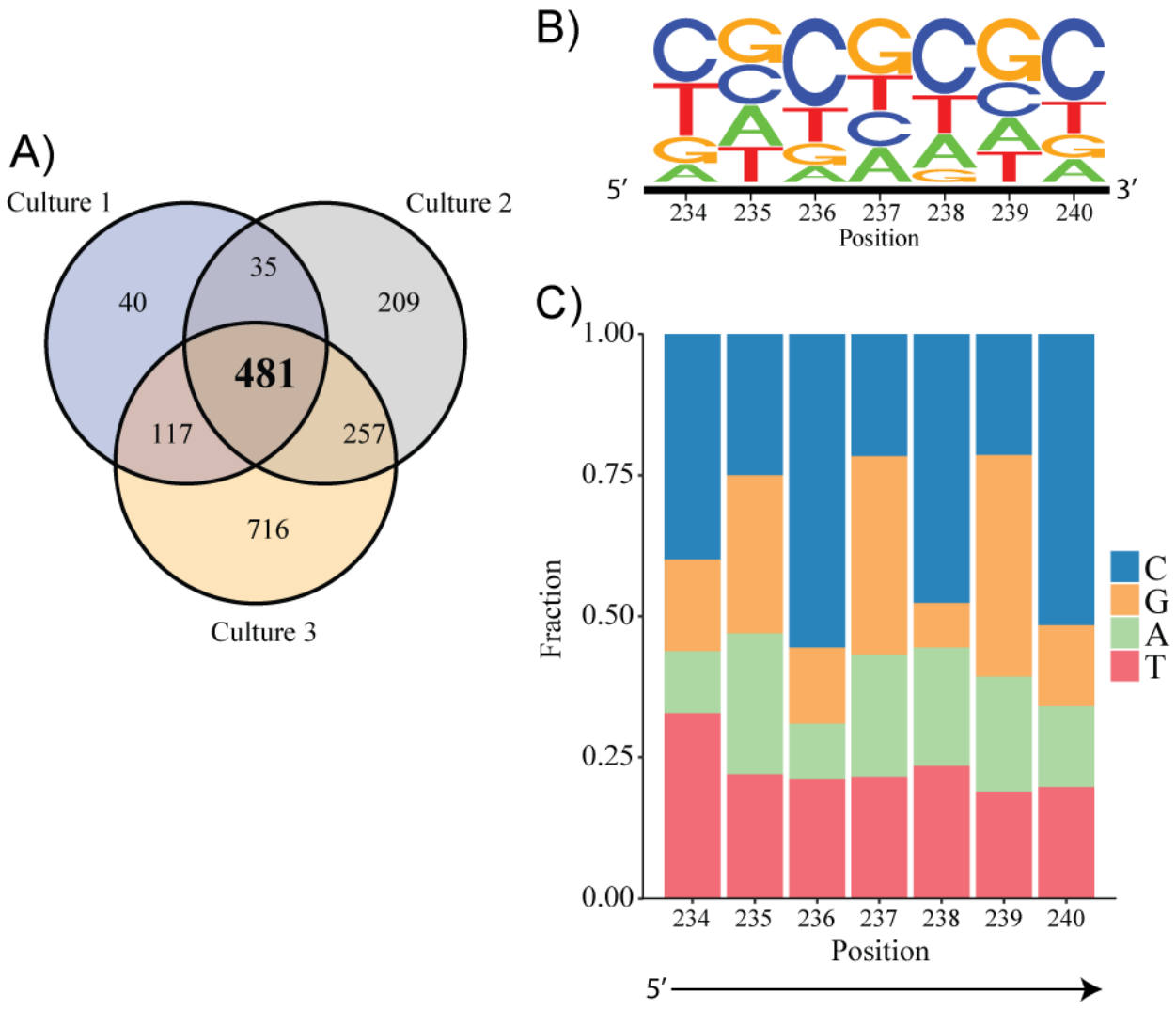


Figure 4.2 Alternating –CG– pattern among shared template mutants. (A) Overlap of template sequences that were present at Round 1 in all three cultures and absent from Round 3 and Round 5 libraries. (B) The 481 template losers shared among all three cultures was analyzed for a common motif. The motif is written starting at the 5' end of the seven nucleotide template of the *ter1* gene. (C) The fraction of nucleotides at each position corresponding to the motif in B.

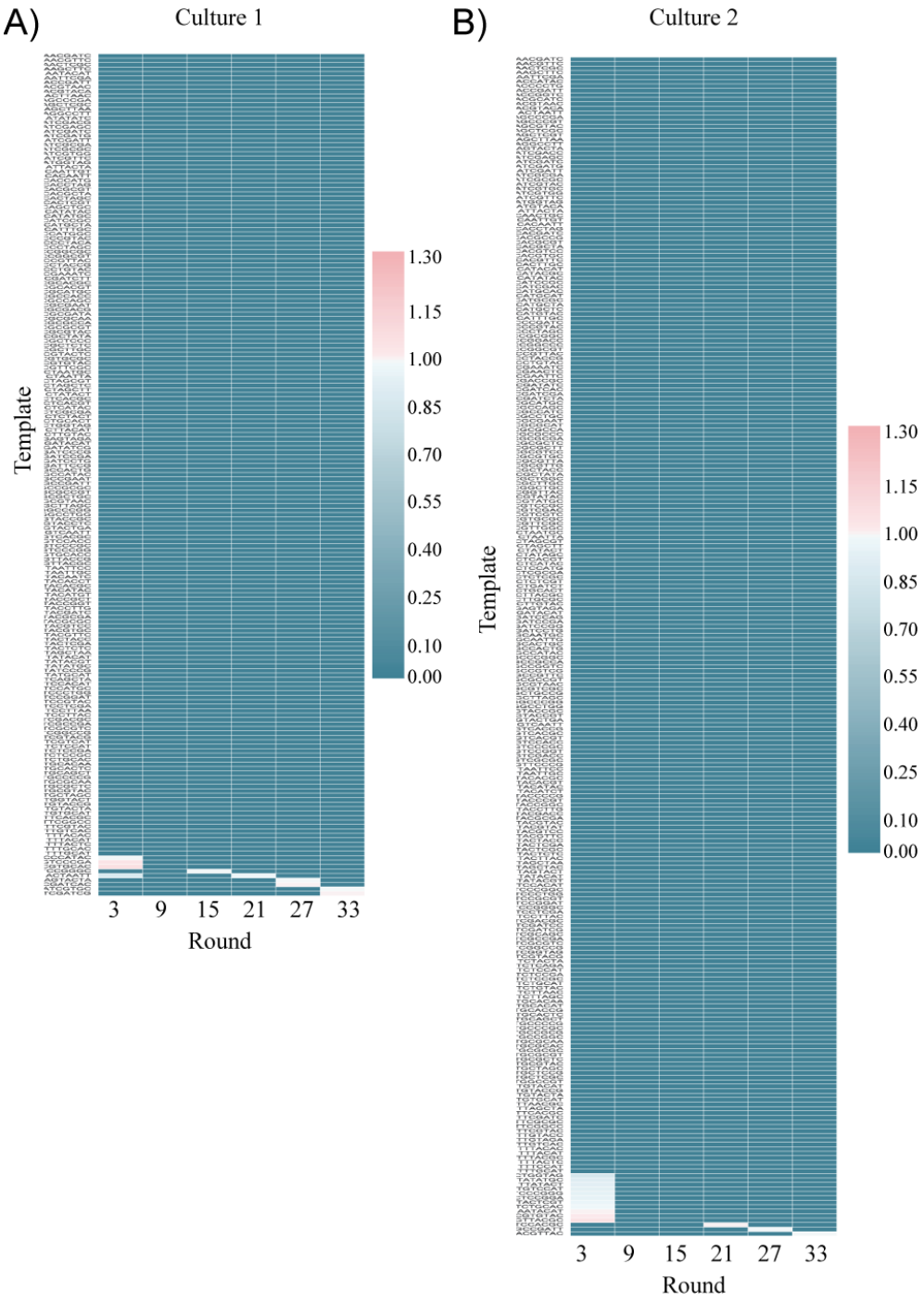


Figure 4.3 Identified 481 template mutants also performed poorly in minimal media. (A, B) The fitness of templates in the minimal media competition. A fitness of 1 is equivalent to wild type. A fitness less than 1 is less fit than wild type. Templates that overlap with the 481 template mutants identified by the log-phase, short time course competition in rich media (Chapter IV) and the Culture 1 of the minimal media experiment of Chapter II (A) or Culture 2 of the minimal media experiment of Chapter II (B) are shown.

IV.5 Discussion

Successful telomere destabilization requires a telomerase RNA template that is used to incorporate mutant repeats into the telomere. Proof of principle experiments in human cancer cell lines have confirmed that expression of mutant templates decreases growth rate and cell viability in a manner dependent on telomerase localization to telomeres and catalytic activity [236, 237]. Template mutations resulted in critically short mutant telomeres, aberrant nuclei indicative of mitotic catastrophe, and chromosome end fusions [239, 240]. Furthermore, co-expression of a mutant telomerase RNA template with a short-interfering RNA (siRNA) against the endogenous template demonstrated a synergistic effect on apoptosis [237]. To define templates that confer the most immediate growth defect, we screened 16,384 *S. pombe* strains with a unique, randomized seven nucleotide *ter1* template in a *ter1*Δ background. Our study is the first comprehensive approach to identifying template mutants that can be used for robust telomere destabilization. We report a 3'-GCGCGC-5' consensus sequence for nucleotides 235-240 among templates that were lost after an average of only 18.1 wild type generations. This consensus sequence differs from the two template mutants used in the human cell line experiments where uracils (U) in the template were replaced with adenines (A) or where an alternating pattern of AU was used [236-238]. Investigation of individual templates from our screen may reveal specific templates that resemble those used in human studies. Further study of template mutants identified by this screen will provide necessary insight into the mechanism of mutant telomere deprotection and DNA Damage response signaling.

IV.6 Future directions

The immediate next steps should focus on reducing the number of templates from this screen for follow-up experiments. A second, short competition could be performed with *ter1*$^{+}$

cells. This experiment would more closely approximate the conditions of treating telomerase positive cancer cells with a mutant-template hTR. Template losers that are shared between the *ter1*$^{-}$ and *ter1*$^{+}$ screens would likely represent RNAs that could block recruitment of wild type telomerase to the chromosome end or elicit a robust DDR at the chromosome end despite interspersed wild type repeats. Various iterations of these competition experiments can be employed to further decrease the number of template candidates. For example, *S. pombe* can be grown in conditions that mimic the tumor intracellular environment including hydrogen peroxide treatment to stimulate oxidative stress, hydroxyurea to cause replication stress, and bleomycin to instigate double strand breaks. In addition to these experiments, template losers should be further analyzed for sequences that would be predicted to generate telomeric repeats unable to bind Taz1 or Pot1. Lastly, this screen can be directly translated into human cancer cell lines with the use of a lentiviral library of randomized hTR templates as begun by a previous lab member, Wuxiang Guan. The screen described here presents a valuable starting point for the identification of mutant telomerase RNA templates that can be used to destabilize telomeres in telomerase-dependent cancer cells. The overlap of template losers between this original screen and proposed variations of the screen will yield a manageable subset of template sequences that can be further characterized for mutant telomere sequences that elicit DNA damage response signaling and subsequent growth arrest in *S. pombe* as a proxy for pre-apoptotic cell cycle arrest in human cancer cells.

Chapter V: Significance, remaining questions, and future experiments

V.1 Contributions to the field of Telomere Biology

The findings reported in this dissertation advance our understanding of telomere evolution, establish a model for repeat addition processivity in *S. pombe,* and provide an entry point for the unbiased identification of telomerase RNA template variants that can be used to destabilize telomeres in cancer cells or screen for telomerase inhibitors. Furthermore, the variant repeat strains identified and characterized in this study can be used in the future investigation of how changes in telomeric Taz1 and Pot1 stoichiometry might affect a) the transduction of telomere length into telomerase recruitment at 3' overhangs of the shortest telomeres and b) the gradation of DNA damage response signaling.

Regarding telomere evolution, researchers have theorized that telomerase evolution preceded the evolution of sequence-specific telomere binding proteins, reasoning that the establishment of a uniform telomeric repeat could have directed the selection of specific DNA recognition motifs [3, 241]. In the present study, protein binding appears to be the major determinant of template success in rich media because the fittest 5nt sequence, 3'-CCCAAU-5' corresponded to telomeric nucleotides known to be important for Taz1 and Pot1 binding (Chapter I.3) without necessarily facilitating telomerase alignment. However, in minimal media, protein binding may play a less decisive role as 3'-UCCUCGG-5' emerged as one of the most frequent templates despite a lack of apparent Taz1 and Pot1 binding sequences in the telomeric repeat.

Here we provide evidence for the first time that variant-template *S. pombe* telomerases can use an alignment register shifted three to four nucleotides downstream of the wild type TER1

alignment region. In contrast to wild type [40], these strains favored translocation sites toward the 3' end of the template, arguing against a fixed structure or interaction as suggested in the *T. thermophila* accordion model. This observation signifies a search for an alignment register that base pairs with the last three nucleotides of the last telomeric repeat and requires considerable flexibility between the telomerase RNA, telomerase catalytic subunit, and telomeric DNA.

At the onset of this research, *S. pombe* and *S. cerevisiae* telomerases were regarded as inherently nonprocessive in contrast to ciliate and vertebrate telomerases. Whether the *S. pombe* enzyme is nonprocessive *in vitro* but processive *in vivo* was unclear. This work demonstrates that the *S. pombe* telomerase can generate nearly perfect repeats *in vivo*. Small changes to the template including a single nucleotide mutation or rearrangement of the nucleotides around the 3'-CCAAU-5' core resulted in robust changes in direct repeat number. Our findings demonstrate that a loss of telomere heterogeneity does not cause a growth disadvantage, excluding degenerate versus direct repeats as a driving evolutionary factor in *S. pombe*. Whereas wild type S. pombe has adopted a complex multi-step catalytic cycle, the variant strains characterized in this study reveal a more simple catalytic cycle that involves a single translocation step. In totality, these findings have informed our view of telomere sequence evolution in *S. pombe* while prompting some important and interesting questions discussed in subsequent sections.

Lastly, this research describes a novel and comprehensive approach to screening for telomere destabilizing template sequences. This is the first application of *in vitro* evolution to identify mutant telomerase RNA templates that cause an abrupt growth defect. The utility of this experiment was made possible by the rapid growth and response to telomere dysfunction in the unicellular telomere model organism, *Schizosaccharomyces pombe*. Importantly, 481 templates were consistently identified among three independent experiments to impair growth after only

three days of in fresh liquid culture, expanding the repertoire of templates that can be used to destabilize telomeres. These templates provide a strong starting point for the further characterization of telomere length maintenance, uncapping, and DNA damage response signaling in mutant template *S. pombe* strains. Analysis of patterns among these 481 templates has also provided new information on the functionally important features of the template that confer a growth defect.

V.2 Does the 3-UCCUCGG-5' template strain grow better in minimal media?

The minimal media competition experiment revealed a slower rate in the loss of templates lacking 3'-CCAAU-5' and an accompanying increase in the frequency of 3'-UCCUCGG-5' template strains by Round 33 compared to the rich media competition (Chapter II.4.5). These findings suggest that 3'-UCCUCGG-5' grows better in minimal than rich media and thereby may represent an example of differences in selection between environmental conditions. However, interpretation of these results is limited by the experimental design. Plasmid libraries used either an auxotrophic marker (*ura4*$^+$) for retention in minimal media or an antibiotic resistance gene (Nourseothricin resistance cassette) for retention in rich media.

A more direct comparison can be made if variant templates are integrated into an *S. pombe* strain at the *ter1*$^+$ genomic locus. Templates should be integrated into a *ura4-* strain for later counterselection against untransformed clones and then crossed with a wild type strain lacking selectable auxotrophic markers. Integrated strains could then be grown in Edinburg Minimal Media (EMM) plus histidine, uracil, adenine, and leucine supplements (minimal media) or YES (rich media). The competitive growth of strains could be assessed in two ways. Firstly, the 3'-UCCUCGG-5' strain could be diluted in a 1:1 ratio with wild type. In this case, template frequency could be monitored by a) using Next Generation Sequencing as described in Chapter

II or b) a *ter1*$^{+}$ probed southern blot of genomic fragments digested with restriction enzymes, MaeIII (Sigma-Aldrich), which uniquely cuts within the wild type 5'-TGTAACC-3' template, and NlaIV (New England Biolabs), which uniquely cuts within the 5'-GGCTCCT-3' template. Secondly, doubling time of each strain could be determined through a time course of 3'-UCCUCGG-5' and wild type strains grown independently.

V.3 With the right TER1 template, can *S. pombe* telomerase become processive?

Understanding the nature of the stop signal in variant template telomerases will inform our model for repeat addition processivity in *S. pombe*. A pause site would favor a human telomerase model whereas a boundary element would be more consistent with the accordion model of *Tetrahymena* (Chapter I.4). The establishment of a mechanism for repeat addition processivity in *S. pombe* would provide an evolutionary link between processive ciliate and vertebrate telomerases. First, the optimal 3' end for repeat addition and increased telomerase activity of variant template telomerases can be confirmed *in vitro* using the direct telomerase activity assay. In this assay, oligonucleotide primers resembling variant telomeric repeats are incubated with immunopurified telomerase and radiolabeled deoxyguanine triphosphate (dGTP). Extension products are then separated by electrophoresis and imaged with a phosphor screen cassette. Second, using the optimized primer, repeat addition processivity can be demonstrated with a pulse-chase experiment. Briefly, telomerase would be incubated with the primer and radiolabeled dGTP in a reaction mixture (pulse) then diluted with an excess of primer and unlabeled dGTP (chase) and extension products would be collected at increasing time intervals. The chase step would sequester unbound telomerase and allow for the quantitation of nucleotides that were added before telomerase units had dissociated [58]. Third, single-molecule Förster resonance energy transfer (smFRET) could be used to determine the mechanical relationship

between the telomeric 3' end and TER1. As done previously with *Tetrahymena* telomerase, oligonucleotides representing the telomeric DNA would be surface-immobilized and coupled to a donor dye [123]. TER1 within the purified telomerase would be coupled to an acceptor dye and the distance between TER1 and the telomeric primer could be assessed as the FRET, or energy transfer, between the two dyes.

V.4 Lessons from variant templates and applications for the innovation of cancer treatment

The research of this dissertation has investigated dysfunctional and functional TER1 templates, which can both be used to exploit telomere biology in the advancement of cancer treatment. One strategy is to instigate a robust DNA damage response (DDR) and subsequent apoptosis in cancer cells by destabilizing their telomeres. To this end, our identification of dysfunctional templates in Chapter IV can be used to decipher features of variant templates that cause telomere uncapping and elicit DDR signaling. Growth defects of select dysfunctional templates should first be confirmed with serial restreaks on plates. Telomere dysfunction can then be assessed by southern blot for the detection of linear telomeres and pulse-field gel electrophoresis for the detection of circularized chromosomes as a consequence of telomere loss and uncapping by Pot1. Additionally, DDR signaling at the telomere can be quantified using fluorescence microscopy and labelled DDR factors as described by Carneiro *et al* [242]. Together, this information can then be used to design hTR template mutants that cause rapid telomere dysfunction and cell death in telomerase-dependent cancers.

A second strategy in cancer therapy is to directly inhibit telomerase catalytic activity. However, the discovery of small molecule inhibitors has been hampered by the lack of relevant high-throughput, cell-based screening assays. Because the ratio of lengths of the 3'overhang to the rest of the telomere is so small, the detection of newly synthesized repeats are obscured by

pre-existing repeats, in turn causing a suboptimal signal to noise ratio for the detection of effects on telomerase activity. In Chapter III, our characterization of functional TER1 template strains revealed a potentially valuable feature. The long stretches of variant telomeric repeats added to a wild type telomere allows for the distinction between newly synthesized repeats and old wild type repeats, while maintaining wild type function, growth, and telomere length. The variant template *ter1*$^{+}$ could then be transformed into *ter1*$^{-}$ cells in the presence or absence of a candidate inhibitor and telomeres could be sequenced or fluorescently labeled to determine the fraction of telomeres elongated with and without treatment. The ability to monitor telomerase-mediated telomere elongation with Fluorescent *In Situ* Hybridization (FISH) labelling of non-wild type repeats has been demonstrated in human cells, demonstrating achievability of this approach [243]. However, this study did not apply this approach in the context of telomerase inhibitors.

V.5 The utility of prospective evolution studies and implications for cancer biology

Prospective or experimental evolution is an over century and half-old concept that has reached renewed vigor with the advent of 21st century sequencing technology. Elena, S.F. and R.E. Lenski pointed out that Charles Darwin noted the importance of experimental evolution in his 1859 book, "On the Origin of Species by Means of Natural Selection" [244]. Charles Darwin wrote, "In looking for the gradations by which an organ in any species has been perfected, we ought to look exclusively to its lineal ancestors; but this is scarcely ever possible, and we are forced in each case to look to species of the same group…" (Darwin, 1859, p. 187) [245]. Since then, microbial model organisms such as bacteria and yeast have frequently been used to study evolution in real time owing to their ability to replicate quickly and reproduce asexually. Next generation sequencing (NGS) has enabled the detection of genetic and epigenetic changes across

the genome at population and single-cell levels and thereby provides a necessary tool for identifying meaningful allele frequency changes over time.

Countless variations on the theme of in-flask evolution can be devised. Richard Lenski's group has been growing *E.coli* since 1988 in a continuous "long-term evolution experiment" to uncover basic principles of evolution [246]. Other studies have started with a yeast strain expressing a library of a specific, mutated noncoding RNA to reveal epistatic interactions and regions important for the RNA's structure, stability, and function [206, 247]. Similar to our study, researchers have also mutated the telomerase RNA template of budding yeast and concluded that its telomerase is specialized for a C-rich template that confers G-rich repeats [219]. Our study of prospective telomere evolution advances this research by introducing next generation sequencing of template mutants and high-throughput telomere cloning and by using fission yeast, a system that shares more telomere conservation with human. This approach has confirmed the selection of C-rich templates and furthermore, allowed the detection of other population-scale changes in template pattern fitness and its impact on telomere function and telomerase repeat synthesis. Additionally, the research described here demonstrates that in-flask evolution can be employed to identify template losers that quickly impair competitive growth. This type of screen will prove to be particularly important as clustered regularly interspaced short palindromic repeats (CRISPR) technology advances the possibility of editing endogenous telomerase RNA template sequences in cancer cells.

Lastly, experimental evolution may provide valuable insight into the evolution of a single tumor. For example, *in vitro* experiments in budding yeast have demonstrated accelerated adaptation rates in tetraploid and aneuploid organisms versus haploids or diploids, supporting the idea that polyploidy can drive tumorigenesis [216]. Like clonally expanding cancer cells,

unicellular organisms acquire mutations that affect their ability to replicate and respond to environmental stimuli and stressors. However, unlike cells within a tumor, unicellular organisms cannot mimic growth in the three-dimensional space of a tumor or account for all of the genetic differences among specific organs or individual patients [248]. Still, the manipulation of growth conditions (e.g. temperature, pH, nutrient source, chemicals, continuous/discontinuous culturing etc.) and starting genetic background (e.g. mutator genotypes, heterogeneous cell populations, gene silencing or deletion etc.) help to approximate tumor microenvironment and cellular heterogeneity. Such experiments can be used to identify genomic loci that are frequently mutated in independent experiments to distinguish driver mutations from passenger mutations, to study epistatic interactions between individual mutations, and to test the response of the mutation landscape to drug treatment [248].

References

1. Cahill, D., B. Connor, and J.P. Carney, *Mechanisms of eukaryotic DNA double strand break repair.* Front Biosci, 2006. **11**: p. 1958-76.

2. Hustedt, N. and D. Durocher, *The control of DNA repair by the cell cycle.* Nat Cell Biol, 2016. **19**(1): p. 1-9.

3. de Lange, T., *How telomeres solve the end-protection problem.* Science, 2009. **326**(5955): p. 948-52.

4. Muller, H., *An analysis of the process of structural change in chromosomes of Drosophila* Journ. of Genetics, 1940. **XL**(1 and 2): p. 1-66.

5. McClintock, B., *The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis.* Proc Natl Acad Sci U S A, 1939. **25**(8): p. 405-16.

6. McClintock, B., *The Stability of Broken Ends of Chromosomes in Zea Mays.* Genetics, 1941. **26**(2): p. 234-82.

7. Muller, H.J., *The Remaking of Chromosomes.* The Collecting Net, 1938. **XIII**(8): p. 181-195.

8. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.

9. Watson, J.D. and F.H. Crick, *Genetical implications of the structure of deoxyribonucleic acid.* Nature, 1953. **171**(4361): p. 964-7.

10. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia Coli.* Proc Natl Acad Sci U S A, 1958. **44**(7): p. 671-82.

11. Taylor, J.H., P.S. Woods, and W.L. Hughes, *The Organization and Duplication of Chromosomes as Revealed by Autoradiographic Studies Using Tritium-Labeled Thymidinee.* Proc Natl Acad Sci U S A, 1957. **43**(1): p. 122-8.

12. Bessman, M.J., et al., *Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction.* J Biol Chem, 1958. **233**(1): p. 171-7.

13. Lehman, I.R., et al., *Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli.* J Biol Chem, 1958. **233**(1): p. 163-70.

14. Alberts, B., et al., *Molecular Biology of the Cell.* 2002, Garland Science;: New York:.

15. Okazaki, R., et al., *Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.* Proc Natl Acad Sci U S A, 1968. **59**(2): p. 598-605.

16. Sugimoto, K., T. Okazaki, and R. Okazaki, *Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages.* Proc Natl Acad Sci U S A, 1968. **60**(4): p. 1356-62.

17. Watson, J.D., *Origin of Concatermic T7 DNA.* Nature, 1972. **239**: p. 197-201.

18. Olovnikov, A.M., *[Principle of marginotomy in template synthesis of polynucleotides]*. Dokl Akad Nauk SSSR, 1971. **201**(6): p. 1496-9.

19. Hayflick, L. and P.S. Moorhead, *The serial cultivation of human diploid cell strains.* Exp Cell Res, 1961. **25**: p. 585-621.

20. Hayflick, L., *The Limited in Vitro Lifetime of Human Diploid Cell Strains.* Exp Cell Res, 1965. **37**: p. 614-36.

21. Shay, J.W. and W.E. Wright, *Hayflick, his limit, and cellular ageing.* Nat Rev Mol Cell Biol, 2000. **1**(1): p. 72-6.

22. Olovnikov, A., *A theory of marginotomy. The incomplete compying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon.* J Theor Biol, 1973. **41**(1): p. 181-190.

23. Podlevsky, J.D., Bley, C.J., Omana, R.V., Qi, X., Chen, J. *The Telomerase Database.* 2007.

24. Nakamura, T.M., et al., *Telomerase catalytic subunit homologs from fission yeast and human.* Science, 1997. **277**(5328): p. 955-9.

25. Nossal, N.G., et al., *Anecdotal, historical and critical commentaries on genetics. Gisela Mosig.* Genetics, 2004. **168**(3): p. 1097-104.

26. Pardue, M.L. and P.G. DeBaryshe, *Drosophila telomeres: A variation on the telomerase theme.* Fly (Austin), 2008. **2**(3): p. 101-10.

27. Blackburn, E.H., C.W. Greider, and J.W. Szostak, *Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging.* Nat Med, 2006. **12**(10): p. 1133-8.

28. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J Mol Biol, 1978. **120**(1): p. 33-53.

29. Szostak, J.W. and E.H. Blackburn, *Cloning yeast telomeres on linear plasmid vectors.* Cell, 1982. **29**(1): p. 245-55.

30. Shampay, J., J.W. Szostak, and E.H. Blackburn, *DNA sequences of telomeres maintained in yeast.* Nature, 1984. **310**(5973): p. 154-7.

31. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.* Cell, 1985. **43**(2 Pt 1): p. 405-13.

32. Greider, C.W. and E.H. Blackburn, *The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity.* Cell, 1987. **51**(6): p. 887-98.

33. Greider, C.W. and E.H. Blackburn, *A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis.* Nature, 1989. **337**(6205): p. 331-7.

34. Blackburn, E.H. and K. Collins, *Telomerase: an RNP enzyme synthesizes DNA.* Cold Spring Harb Perspect Biol, 2011. **3**(5): p. pii:a003558.

35. Lingner, J., et al., *Reverse transcriptase motifs in the catalytic subunit of telomerase.* Science, 1997. **276**(5312): p. 561-7.

36. Collins, K. and L. Gandhi, *The reverse transcriptase component of the Tetrahymena telomerase ribonucleoprotein complex.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8485-90.

37. Egan, E.D. and K. Collins, *Biogenesis of telomerase ribonucleoproteins.* RNA, 2012. **18**(10): p. 1747-59.

38. Chen, J.L., M.A. Blasco, and C.W. Greider, *Secondary structure of vertebrate telomerase RNA.* Cell, 2000. **100**(5): p. 503-14.

39. Feng, J., et al., *The RNA component of human telomerase.* Science, 1995. **269**(5228): p. 1236-41.

40. Leonardi, J., et al., *TER1, the RNA subunit of fission yeast telomerase.* Nat Struct Mol Biol, 2008. **15**(1): p. 26-33.

41. Podlevsky, J.D. and J.J. Chen, *Evolutionary perspectives of telomerase RNA structure and function.* RNA Biol, 2016. **13**(8): p. 720-32.

42. McCormick-Graham, M. and D.P. Romero, *Ciliate telomerase RNA structural features.* Nucleic Acids Res, 1995. **23**(7): p. 1091-7.

43. Qi, X., et al., *The common ancestral core of vertebrate and fungal telomerase RNAs.* Nucleic Acids Res, 2013. **41**(1): p. 450-62.

44. Webb, C.J. and V.A. Zakian, *Telomerase RNA stem terminus element affects template boundary element function, telomere sequence, and shelterin binding.* Proc Natl Acad Sci U S A, 2015. **112**(36): p. 11312-7.

45. Webb, C.J. and V.A. Zakian, *Identification and characterization of the Schizosaccharomyces pombe TER1 telomerase RNA.* Nat Struct Mol Biol, 2008. **15**(1): p. 34-42.

46. Box, J.A., et al., *A flexible template boundary element in the RNA subunit of fission yeast telomerase.* J Biol Chem, 2008. **283**(35): p. 24224-33.

47. Webb, C.J. and V.A. Zakian, *Telomerase RNA is more than a DNA template.* RNA Biol, 2016. **13**(8): p. 683-9.

48. Meyne, J., R.L. Ratliff, and R.K. Moyzis, *Conservation of the human telomere sequence (TTAGGG)n among vertebrates.* Proc Natl Acad Sci U S A, 1989. **86**(18): p. 7049-53.

49. Moyzis, R.K., et al., *A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes.* Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6622-6.

50. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.

51. de Lange, T., et al., *Structure and variability of human chromosome ends.* Mol Cell Biol, 1990. **10**(2): p. 518-27.

52. Zhao, Y., et al., *Quantitative telomeric overhang determination using a double-strand specific nuclease.* Nucleic Acids Res, 2008. **36**(3): p. e14.

53. Chai, W., et al., *Human telomeres have different overhang sizes at leading versus lagging strands.* Mol Cell, 2006. **21**(3): p. 427-35.

54. Makarov, V.L., Y. Hirose, and J.P. Langmore, *Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening.* Cell, 1997. **88**(5): p. 657-66.

55. Wright, W.E., et al., *Normal human chromosomes have long G-rich telomeric overhangs at one end.* Genes Dev, 1997. **11**(21): p. 2801-9.

56. Kipling, D. and H.J. Cooke, *Hypervariable ultra-long telomeres in mice.* Nature, 1990. **347**(6291): p. 400-2.

57. Morin, G.B., *The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats.* Cell, 1989. **59**(3): p. 521-9.

58. Greider, C.W., *Telomerase is processive.* Mol Cell Biol, 1991. **11**(9): p. 4572-80.

59. McEachern, M.J. and E.H. Blackburn, *A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3453-7.

60. Sugawara, N.F., *DNA Sequences at the Telomeres of the Fission Yeast S. pombe*. 1988, Harvard University: Cambridge, MA.

61. Trujillo, K.M., J.T. Bunch, and P. Baumann, *Extended DNA binding site in Pot1 broadens sequence specificity to allow recognition of heterogeneous fission yeast telomeres.* J Biol Chem, 2005. **280**(10): p. 9119-28.

62. Nakamura, T.M., J.P. Cooper, and T.R. Cech, *Two modes of survival of fission yeast without telomerase.* Science, 1998. **282**(5388): p. 493-6.

63. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.

64. Cooper, J.P., et al., *Regulation of telomere length and function by a Myb-domain protein in fission yeast.* Nature, 1997. **385**(6618): p. 744-7.

65. Bianchi, A., et al., *TRF1 is a dimer and bends telomeric DNA.* EMBO J, 1997. **16**(7): p. 1785-94.

66. Bilaud, T., et al., *Telomeric localization of TRF2, a novel human telobox protein.* Nat Genet, 1997. **17**(2): p. 236-9.

67. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.

68. Chen, Y., et al., *A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms.* Nat Struct Mol Biol, 2011. **18**(2): p. 213-21.

69. Chikashige, Y. and Y. Hiraoka, *Telomere binding of the Rap1 protein is required for meiosis in fission yeast.* Curr Biol, 2001. **11**(20): p. 1618-23.

70. Kanoh, J. and F. Ishikawa, *spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast.* Curr Biol, 2001. **11**(20): p. 1624-30.

71. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.

72. Lei, M., P. Baumann, and T.R. Cech, *Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of Schizosaccharomyces pombe.* Biochemistry, 2002. **41**(49): p. 14560-8.

73. Croy, J.E., E.R. Podell, and D.S. Wuttke, *A new model for Schizosaccharomyces pombe telomere recognition: the telomeric single-stranded DNA-binding activity of Pot11-389.* J Mol Biol, 2006. **361**(1): p. 80-93.

74. Dickey, T.H., M.A. McKercher, and D.S. Wuttke, *Nonspecific recognition is achieved in Pot1pC through the use of multiple binding modes.* Structure, 2013. **21**(1): p. 121-32.

75. Dickey, T.H. and D.S. Wuttke, *The telomeric protein Pot1 from Schizosaccharomyces pombe binds ssDNA in two modes with differing 3' end availability.* Nucleic Acids Res, 2014. **42**(15): p. 9656-65.

76. Ye, J.Z., et al., *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex.* Genes Dev, 2004. **18**(14): p. 1649-54.

77. Kim, S.H., P. Kaminker, and J. Campisi, *TIN2, a new regulator of telomere length in human cells.* Nat Genet, 1999. **23**(4): p. 405-12.

78. O'Connor, M.S., et al., *A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11874-9.

79. Ye, J.Z., et al., *TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres.* J Biol Chem, 2004. **279**(45): p. 47264-71.

80. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.

81. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.

82. Nandakumar, J. and T.R. Cech, *DNA-induced dimerization of the single-stranded DNA binding telomeric protein Pot1 from Schizosaccharomyces pombe.* Nucleic Acids Res, 2012. **40**(1): p. 235-44.

83. Altschuler SE, D.T., Wuttke DS, *Schizosaccharomyces pombe Protection of Telomeres 1 Utilizes Alternate Binding Modes to Accomodate Different Telomeric Sequences.* Biochemistry, 2011. **50**(35): p. 7503-13.

84. Altschuler, S.E., T.H. Dickey, and D.S. Wuttke, *Schizosaccharomyces pombe protection of telomeres 1 utilizes alternate binding modes to accommodate different telomeric sequences.* Biochemistry, 2011. **50**(35): p. 7503-13.

85. Maciejowski, J. and T. de Lange, *Telomeres in cancer: tumour suppression and genome instability.* Nat Rev Mol Cell Biol, 2017. **18**(3): p. 175-186.

86. Doksani, Y. and T. de Lange, *The role of double-strand break repair pathways at functional and dysfunctional telomeres.* Cold Spring Harb Perspect Biol, 2014. **6**(12): p. a016576.

87. Ferreira, M.G. and J.P. Cooper, *The fission yeast Taz1 protein protects chromosomes from Ku-dependent end-to-end fusions.* Mol Cell, 2001. **7**(1): p. 55-63.

88. Ferreira, M.G. and J.P. Cooper, *Two modes of DNA double-strand break repair are reciprocally regulated through the fission yeast cell cycle.* Genes Dev, 2004. **18**(18): p. 2249-54.

89. van Steensel, B., A. Smogorzewska, and T. de Lange, *TRF2 protects human telomeres from end-to-end fusions.* Cell, 1998. **92**(3): p. 401-13.

90. Smogorzewska, A., et al., *DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2.* Curr Biol, 2002. **12**(19): p. 1635-44.

91. Celli, G.B. and T. de Lange, *DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion.* Nat Cell Biol, 2005. **7**(7): p. 712-8.

92. Martinez, P., et al., *Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice.* Genes Dev, 2009. **23**(17): p. 2060-75.

93. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang.* EMBO J, 2001. **20**(19): p. 5532-40.

94. Doksani, Y., et al., *Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation.* Cell, 2013. **155**(2): p. 345-56.

95. Bandaria, J.N., et al., *Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin.* Cell, 2016. **164**(4): p. 735-46.

96. Timashev, L.A., et al., *The DDR at telomeres lacking intact shelterin does not require substantial chromatin decompaction.* Genes Dev, 2017. **31**(6): p. 578-589.

97. Vancevska, A., et al., *The telomeric DNA damage response occurs in the absence of chromatin decompaction.* Genes Dev, 2017. **31**(6): p. 567-577.

98. Fabregat, A., et al., *The Reactome pathway Knowledgebase.* Nucleic Acids Res, 2016. **44**(D1): p. D481-7.

99. Wang, X. and P. Baumann, *Chromosome fusions following telomere loss are mediated by single-strand annealing.* Mol Cell, 2008. **31**(4): p. 463-73.

100. Capper, R., et al., *The nature of telomere fusion and a definition of the critical telomere length in human cells.* Genes Dev, 2007. **21**(19): p. 2495-508.

101. He, H., et al., *POT1b protects telomeres from end-to-end chromosomal fusions and aberrant homologous recombination.* EMBO J, 2006. **25**(21): p. 5180-90.

102. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. **126**(1): p. 49-62.

103. Schmidt, J.C. and T.R. Cech, *Human telomerase: biogenesis, trafficking, recruitment, and activation.* Genes Dev, 2015. **29**(11): p. 1095-105.

104. Mozdy, A.D. and T.R. Cech, *Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency.* RNA, 2006. **12**(9): p. 1721-37.

105. Xi, L. and T.R. Cech, *Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT.* Nucleic Acids Res, 2014. **42**(13): p. 8565-77.

106. Liu, Y., et al., *Preferential maintenance of critically short telomeres in mammalian cells heterozygous for mTert.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3597-602.

107. Teixeira, M.T., et al., *Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states.* Cell, 2004. **117**(3): p. 323-35.

108. Zhao, Y., et al., *Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells.* Cell, 2009. **138**(3): p. 463-75.

109. Zhao, Y., et al., *Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions.* Mol Cell, 2011. **42**(3): p. 297-307.

110. Schmidt, J.C., A.J. Zaug, and T.R. Cech, *Live Cell Imaging Reveals the Dynamics of Telomerase Recruitment to Telomeres.* Cell, 2016. **166**(5): p. 1188-1197 e9.

111. Cesare, A.J. and J. Karlseder, *A three-state model of telomere control over human proliferative boundaries.* Curr Opin Cell Biol, 2012. **24**(6): p. 731-8.

112. Pan, L., et al., *Minishelterins separate telomere length regulation and end protection in fission yeast.* Genes Dev, 2015. **29**(11): p. 1164-74.

113. Hu, X., et al., *Multi-step coordination of telomerase recruitment in fission yeast through two coupled telomere-telomerase interfaces.* Elife, 2016. **5**.

114. Flory, M.R., et al., *An SMC-domain protein in fission yeast links telomeres to the meiotic centrosome.* Mol Cell, 2004. **16**(4): p. 619-30.

115. Moser, B.A., O.N. Raguimova, and T.M. Nakamura, *Ccq1-Tpz1TPP1 interaction facilitates telomerase and SHREC association with telomeres in fission yeast.* Mol Biol Cell, 2015. **26**(21): p. 3857-66.

116. Tomita, K. and J.P. Cooper, *Fission yeast Ccq1 is telomerase recruiter and local checkpoint controller.* Genes Dev, 2008. **22**(24): p. 3461-74.

117. Moser, B.A., et al., *Tel1ATM and Rad3ATR kinases promote Ccq1-Est1 interaction to maintain telomeres in fission yeast.* Nat Struct Mol Biol, 2011. **18**(12): p. 1408-13.

118. Yamazaki, H., Y. Tarumoto, and F. Ishikawa, *Tel1(ATM) and Rad3(ATR) phosphorylate the telomere protein Ccq1 to recruit telomerase and elongate telomeres in fission yeast.* Genes Dev, 2012. **26**(3): p. 241-6.

119. Chang, Y.T., B.A. Moser, and T.M. Nakamura, *Fission yeast shelterin regulates DNA polymerases and Rad3(ATR) kinase to limit telomere extension.* PLoS Genet, 2013. **9**(11): p. e1003936.

120. Beernink, H.T., et al., *Telomere maintenance in fission yeast requires an Est1 ortholog.* Curr Biol, 2003. **13**(7): p. 575-80.

121. Webb, C.J. and V.A. Zakian, *Schizosaccharomyces pombe Ccq1 and TER1 bind the 14-3-3-like domain of Est1, which promotes and stabilizes telomerase-telomere association.* Genes Dev, 2012. **26**(1): p. 82-91.

122. Armstrong, C.A., et al., *Telomerase activation after recruitment in fission yeast.* Curr Biol, 2014. **24**(17): p. 2006-11.

123. Berman, A.J., et al., *The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis.* Nat Struct Mol Biol, 2011. **18**(12): p. 1371-5.

124. Nandakumar, J., et al., *The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity.* Nature, 2012. **492**(7428): p. 285-9.

125. Prescott, J. and E.H. Blackburn, *Telomerase RNA mutations in Saccharomyces cerevisiae alter telomerase action and reveal nonprocessivity in vivo and in vitro.* Genes Dev, 1997. **11**(4): p. 528-40.

126. Parks, J.W. and M.D. Stone, *Coordinated DNA dynamics during the human telomerase catalytic cycle.* Nat Commun, 2014. **5**: p. 4146.

127. Brown, A.F., et al., *A self-regulating template in human telomerase.* Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11311-6.

128. Wang F, P.E., Zaug AJ, Yang Y, Baciu P, Cech TR, Lei M, *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-510.

129. Gramatges, M.M., et al., *A homozygous telomerase T-motif variant resulting in markedly reduced repeat addition processivity in siblings with Hoyeraal Hreidarsson syndrome.* Blood, 2013. **121**(18): p. 3586-93.

130. Alder, J.K., et al., *Ancestral mutation in telomerase causes defects in repeat addition processivity and manifests as familial pulmonary fibrosis.* PLoS Genet, 2011. **7**(3): p. e1001352.

131. Pascolo, E., et al., *Mechanism of human telomerase inhibition by BIBR1532, a synthetic, non-nucleosidic drug candidate.* J Biol Chem, 2002. **277**(18): p. 15566-72.

132. Lai, C.K., M.C. Miller, and K. Collins, *Roles for RNA in telomerase nucleotide and repeat addition processivity.* Mol Cell, 2003. **11**(6): p. 1673-83.

133. Autexier, C. and C.W. Greider, *Boundary elements of the Tetrahymena telomerase RNA template and alignment domains.* Genes Dev, 1995. **9**(18): p. 2227-39.

134. Tzfati, Y., et al., *Template boundary in a yeast telomerase specified by RNA structure.* Science, 2000. **288**(5467): p. 863-7.

135. Seto, A.G., et al., *A template-proximal RNA paired element contributes to Saccharomyces cerevisiae telomerase activity.* RNA, 2003. **9**(11): p. 1323-32.

136. Lai, C.K., M.C. Miller, and K. Collins, *Template boundary definition in Tetrahymena telomerase.* Genes Dev, 2002. **16**(4): p. 415-20.

137. Miller, M.C., J.K. Liu, and K. Collins, *Template definition by Tetrahymena telomerase reverse transcriptase.* EMBO J, 2000. **19**(16): p. 4412-22.

138. Ares, M., Jr. and K. Chakrabarti, *Stuttering against marginotomy.* Nat Struct Mol Biol, 2008. **15**(1): p. 18-9.

139. Armanios, M. and E.H. Blackburn, *The telomere syndromes.* Nat Rev Genet, 2012. **13**(10): p. 693-704.

140. Buseman, C.M., W.E. Wright, and J.W. Shay, *Is telomerase a viable target in cancer?* Mutat Res, 2012. **730**(1-2): p. 90-7.

141. Shay, J.W. and W.E. Wright, *Role of telomeres and telomerase in cancer.* Semin Cancer Biol, 2011. **21**(6): p. 349-53.

142. Bodnar, A.G., et al., *Extension of life-span by introduction of telomerase into normal human cells.* Science, 1998. **279**(5349): p. 349-52.

143. d'Adda di Fagagna, F., et al., *A DNA damage checkpoint response in telomere-initiated senescence.* Nature, 2003. **426**(6963): p. 194-8.

144. Takai, H., A. Smogorzewska, and T. de Lange, *DNA damage foci at dysfunctional telomeres.* Curr Biol, 2003. **13**(17): p. 1549-56.

145. Roger, L., et al., *Extensive telomere erosion in the initiation of colorectal adenomas and its association with chromosomal instability.* J Natl Cancer Inst, 2013. **105**(16): p. 1202-11.

146. Riboni, R., et al., *Telomeric fusions in cultured human fibroblasts as a source of genomic instability.* Cancer Genet Cytogenet, 1997. **95**(2): p. 130-6.

147. Artandi, S.E., et al., *Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice.* Nature, 2000. **406**(6796): p. 641-5.

148. Maciejowski, J., et al., *Chromothripsis and Kataegis Induced by Telomere Crisis.* Cell, 2015. **163**(7): p. 1641-54.

149. O'Hagan, R.C., et al., *Telomere dysfunction provokes regional amplification and deletion in cancer genomes.* Cancer Cell, 2002. **2**(2): p. 149-55.

150. Shain, A.H., et al., *The Genetic Evolution of Melanoma from Precursor Lesions.* N Engl J Med, 2015. **373**(20): p. 1926-36.

151. Nault, J.C., et al., *High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions.* Nat Commun, 2013. **4**: p. 2218.

152. Wang, N., et al., *TERT promoter mutation as an early genetic event activating telomerase in follicular thyroid adenoma (FTA) and atypical FTA.* Cancer, 2014. **120**(19): p. 2965-79.

153. Peifer, M., et al., *Telomerase activation by genomic rearrangements in high-risk neuroblastoma.* Nature, 2015. **526**(7575): p. 700-4.

154. Valentijn, L.J., et al., *TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors.* Nat Genet, 2015. **47**(12): p. 1411-4.

155. Sinha-Datta, U., et al., *Transcriptional activation of hTERT through the NF-kappaB pathway in HTLV-I-transformed cells.* Blood, 2004. **104**(8): p. 2523-31.

156. Straat, K., et al., *Activation of telomerase by human cytomegalovirus.* J Natl Cancer Inst, 2009. **101**(7): p. 488-97.

157. Liu, X., et al., *HPV E7 contributes to the telomerase activity of immortalized and tumorigenic cells and augments E6-induced hTERT promoter function.* Virology, 2008. **375**(2): p. 611-23.

158. Bellon, M. and C. Nicot, *Regulation of telomerase and telomeres: human tumor viruses take control.* J Natl Cancer Inst, 2008. **100**(2): p. 98-108.

159. Hoffmeyer, K., et al., *Wnt/beta-catenin signaling regulates telomerase in stem cells and cancer cells.* Science, 2012. **336**(6088): p. 1549-54.

160. Zhang, Y., et al., *Human telomerase reverse transcriptase (hTERT) is a novel target of the Wnt/beta-catenin pathway in human cancer.* J Biol Chem, 2012. **287**(39): p. 32494-511.

161. Liu, C., et al., *The telomerase reverse transcriptase (hTERT) gene is a direct target of the histone methyltransferase SMYD3.* Cancer Res, 2007. **67**(6): p. 2626-31.

162. Li, H., T.H. Lee, and H. Avraham, *A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer.* J Biol Chem, 2002. **277**(23): p. 20965-73.

163. Zhao, Y., et al., *Dual roles of c-Myc in the regulation of hTERT gene.* Nucleic Acids Res, 2014. **42**(16): p. 10385-98.

164. Stern, J.L., et al., *Mutation of the TERT promoter, switch to active chromatin, and monoallelic TERT expression in multiple cancers.* Genes Dev, 2015. **29**(21): p. 2219-24.

165. Huang, F.W., et al., *Highly recurrent TERT promoter mutations in human melanoma.* Science, 2013. **339**(6122): p. 957-9.

166. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma.* Science, 2013. **339**(6122): p. 959-61.

167. Weinhold, N., et al., *Genome-wide analysis of noncoding regulatory mutations in cancer.* Nat Genet, 2014. **46**(11): p. 1160-5.

168. Heidenreich, B. and R. Kumar, *TERT promoter mutations in telomere biology.* Mutat Res, 2017. **771**: p. 15-31.

169. Liu, T., X. Yuan, and D. Xu, *Cancer-Specific Telomerase Reverse Transcriptase (TERT) Promoter Mutations: Biological and Clinical Implications.* Genes (Basel), 2016. **7**(7).

170. Chiba, K., et al., *Cancer-associated TERT promoter mutations abrogate telomerase silencing.* Elife, 2015. **4**.

171. Heidenreich, B., et al., *TERT promoter mutations and telomere length in adult malignant gliomas and recurrences.* Oncotarget, 2015. **6**(12): p. 10617-33.

172. Borah, S., et al., *Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer.* Science, 2015. **347**(6225): p. 1006-10.

173. Bell, R.J., et al., *Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer.* Science, 2015. **348**(6238): p. 1036-9.

174. Huang, F.W., et al., *TERT promoter mutations and monoallelic activation of TERT in cancer.* Oncogenesis, 2015. **4**: p. e176.

175. Aeby, E., et al., *Peroxiredoxin 1 Protects Telomeres from Oxidative Damage and Preserves Telomeric DNA for Extension by Telomerase.* Cell Rep, 2016. **17**(12): p. 3107-3114.

176. Ludlow, A.T., et al., *Telomeres shorten in response to oxidative stress in mouse skeletal muscle fibers.* J Gerontol A Biol Sci Med Sci, 2014. **69**(7): p. 821-30.

177. von Zglinicki, T., *Oxidative stress shortens telomeres.* Trends Biochem Sci, 2002. **27**(7): p. 339-44.

178. Fouquerel, E., et al., *Oxidative guanine base damage regulates human telomerase activity.* Nat Struct Mol Biol, 2016. **23**(12): p. 1092-1100.

179. Indran, I.R., M.P. Hande, and S. Pervaiz, *hTERT overexpression alleviates intracellular ROS production, improves mitochondrial function, and inhibits ROS-mediated apoptosis in cancer cells.* Cancer Res, 2011. **71**(1): p. 266-76.

180. Luiten, R.M., et al., *Ectopic hTERT expression extends the life span of human CD4+ helper and regulatory T-cell clones and confers resistance to oxidative stress-induced apoptosis.* Blood, 2003. **101**(11): p. 4512-9.

181. Ahmed, S., et al., *Telomerase does not counteract telomere shortening but protects mitochondrial function under oxidative stress.* J Cell Sci, 2008. **121**(Pt 7): p. 1046-53.

182. Dudognon, C., et al., *Death receptor signaling regulatory function for telomerase: hTERT abolishes TRAIL-induced apoptosis, independently of telomere maintenance.* Oncogene, 2004. **23**(45): p. 7469-74.

183. Lee, J., et al., *TERT promotes cellular and organismal survival independently of telomerase activity.* Oncogene, 2008. **27**(26): p. 3754-60.

184. Cao, Y., et al., *TERT regulates cell survival independent of telomerase enzymatic activity.* Oncogene, 2002. **21**(20): p. 3130-8.

185. Fougere, B., et al., *Chronic Inflammation: Accelerator of Biological Aging.* J Gerontol A Biol Sci Med Sci, 2016.

186. Zhang, J., et al., *Ageing and the telomere connection: An intimate relationship with inflammation.* Ageing Res Rev, 2016. **25**: p. 55-69.

187. Akiyama, M., et al., *Cytokines modulate telomerase activity in a human multiple myeloma cell line.* Cancer Res, 2002. **62**(13): p. 3876-82.

188. Teo, H., et al., *Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression.* Nat Cell Biol, 2010. **12**(8): p. 758-67.

189. Ghosh, A., et al., *Telomerase directly regulates NF-kappaB-dependent transcription.* Nat Cell Biol, 2012. **14**(12): p. 1270-81.

190. Yin, L., A.K. Hubbard, and C. Giardina, *NF-kappa B regulates transcription of the mouse telomerase catalytic subunit.* J Biol Chem, 2000. **275**(47): p. 36671-5.

191. Robinson, N.J. and W.P. Schiemann, *Means to the ends: The role of telomeres and telomere processing machinery in metastasis.* Biochim Biophys Acta, 2016. **1866**(2): p. 320-329.

192. Park, J.I., et al., *Telomerase modulates Wnt signalling by association with target gene chromatin.* Nature, 2009. **460**(7251): p. 66-72.

193. Smith, L.L., H.A. Coller, and J.M. Roberts, *Telomerase modulates expression of growth-controlling genes and enhances cell proliferation.* Nat Cell Biol, 2003. **5**(5): p. 474-9.

194. Zhou, L., et al., *Telomerase reverse transcriptase activates the expression of vascular endothelial growth factor independent of telomerase activity.* Biochem Biophys Res Commun, 2009. **386**(4): p. 739-43.

195. Qin, Y., et al., *An hTERT/ZEB1 complex directly regulates E-cadherin to promote epithelial-to-mesenchymal transition (EMT) in colorectal cancer.* Oncotarget, 2016. **7**(1): p. 351-61.

196. Liu, Z., et al., *Telomerase reverse transcriptase promotes epithelial-mesenchymal transition and stem cell-like traits in cancer cells.* Oncogene, 2013. **32**(36): p. 4203-13.

197. Wu, S., et al., *Telomerase reverse transcriptase gene promoter mutations help discern the origin of urogenital tumors: a genomic and molecular study.* Eur Urol, 2014. **65**(2): p. 274-7.

198. Bagheri, S., et al., *Genes and pathways downstream of telomerase in melanoma metastasis.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11306-11.

199. Agrawal, A., S. Dang, and R. Gabrani, *Recent patents on anti-telomerase cancer therapy.* Recent Pat Anticancer Drug Discov, 2012. **7**(1): p. 102-17.

200. Zanetti, M., *A second chance for telomerase reverse transcriptase in anticancer immunotherapy.* Nat Rev Clin Oncol, 2017. **14**(2): p. 115-128.

201. Viergever, R.F. and D. Ghersi, *The ClinicalTrials.gov results database.* N Engl J Med, 2011. **364**(22): p. 2169-70; author reply 2170.

202. Institute, N.C. *NCI Drug Dictionary*. May 04, 2017]; Available from: https://www.cancer.gov/publications/dictionaries/cancer-drug?search=p53.

203. Tang, W., et al., *Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes.* Nature, 2012. **484**(7393): p. 260-4.

204. Ewing, B. and P. Green, *Base-calling of automated sequencer traces using phred. II. Error probabilities.* Genome Res, 1998. **8**(3): p. 186-94.

205. Illumina, I. *Quality Scores for Next-Generation Sequencing Assessing sequencing accuracy using Phred quality scoring.* Technical Note: Sequencing 2011 31 October 2011 [cited 2017 25 March]; Available from: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf

206. Li, C., et al., *The fitness landscape of a tRNA gene*. Science, 2016. **352**(6287): p. 837-40.

207. Wickham, H. and W. Chang. *A box and whiskers plot (in the style of Tukey)*. ggplot2 [cited 2017 May 03, 2017]; Available from: http://ggplot2.tidyverse.org/reference/geom_boxplot.html.

208. Lee, M., et al., *Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes.* Nucleic Acids Res, 2014. **42**(3): p. 1733-46.

209. Ye, A.J. and D.P. Romero, *A unique pause pattern during telomere addition by the error-prone telomerase from the ciliate Paramecium tetraurelia.* Gene, 2002. **294**(1-2): p. 205-13.

210. Fulneckova, J., et al., *A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes.* Genome Biol Evol, 2013. **5**(3): p. 468-83.

211. Kuprys PV, D.S., Hauer TM, Meltser M, Tzfati Y, Kirk KE, *Identification of telomerase RNAs from filamentous fungi reveals conservation with vertebrates and yeasts.* PLoS One, 2013. **8**(3): p. e58661.

212. Gunisova, S., et al., *Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements.* RNA, 2009. **15**(4): p. 546-59.

213. Zhong Z, S.L., Kaplan S, de Lange T, *A mammalian factor that binds telomeric TTAGGG repeats in vitro.* Mol Cell Biol, 1992. **12**(11): p. 4834-43.

214. Chong, L., et al., *A human telomeric protein.* Science, 1995. **270**(5242): p. 1663-7.

215. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.

216. Selmecki, A.M., et al., *Polyploidy can drive rapid adaptation in yeast.* Nature, 2015. **519**(7543): p. 349-52.

217. Petersen, J. and P. Russell, *Growth and the Environment of Schizosaccharomyces pombe.* Cold Spring Harb Protoc, 2016. **2016**(3): p. pdb top079764.

218. Kannan, R., et al., *Diverse mechanisms for spliceosome-mediated 3' end processing of telomerase RNA.* Nat Commun, 2015. **6**: p. 6104.

219. Forstemann, K., et al., *Yeast telomerase is specialized for C/A-rich RNA templates.* Nucleic Acids Res, 2003. **31**(6): p. 1646-55.

220. Nurse, P., *Genetic control of cell size at cell division in yeast.* Nature, 1975. **256**(5518): p. 547-51.

221. Egel, R., *Physiological aspects of conjugation in fission yeast.* Planta, 1971. **98**(1): p. 89-96.

222. Sepsiova, R., et al., *Evolution of Telomeres in Schizosaccharomyces pombe and Its Possible Relationship to the Diversification of Telomere Binding Proteins.* PLoS One, 2016. **11**(4): p. e0154225.

223. Steinberg-Neifach, O. and N.F. Lue, *Telomere DNA recognition in Saccharomycotina yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites.* Front Genet, 2015. **6**: p. 162.

224. Wellinger, R.J. and D. Sen, *The DNA structures at the ends of eukaryotic chromosomes.* Eur J Cancer, 1997. **33**(5): p. 735-49.

225. Church, G.M. and W. Gilbert, *Genomic sequencing.* Proc Natl Acad Sci U S A, 1984. **81**(7): p. 1991-5.

226. Konig, P. and D. Rhodes, *Recognition of telomeric DNA.* Trends Biochem Sci, 1997. **22**(2): p. 43-7.

227. Hiraoka, Y., E. Henderson, and E.H. Blackburn, *Not so peculiar: fission yeast telomere repeats.* Trends Biochem Sci, 1998. **23**(4): p. 126.

228. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

229. Drosopoulos, W.C., R. Direnzo, and V.R. Prasad, *Human telomerase RNA template sequence is a determinant of telomere repeat extension rate.* J Biol Chem, 2005. **280**(38): p. 32801-10.

230. Henning, K.A., et al., *Humanizing the yeast telomerase template.* Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5667-71.

231. Qi X, X.M., Brown AF, Bley CJ, Podlevsky JD, Chen JJ, *RNA/DNA hybrid binding affinity determines telomerase template-translocation efficiency.* EMBO J, 2012. **31**(1): p. 150-61.

232. Bryan, T.M., K.J. Goodrich, and T.R. Cech, *A mutant of Tetrahymena telomerase reverse transcriptase with increased processivity.* J Biol Chem, 2000. **275**(31): p. 24199-207.

233. Rivera, M.A. and E.H. Blackburn, *Processive utilization of the human telomerase template: lack of a requirement for template switching.* J Biol Chem, 2004. **279**(51): p. 53770-81.

234. Yu, G.L., et al., *In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs.* Nature, 1990. **344**(6262): p. 126-32.

235. Artandi, S.E., *Complex roles for telomeres and telomerase in breast carcinogenesis.* Breast Cancer Res, 2003. **5**(1): p. 37-41.

236. Goldkorn, A. and E.H. Blackburn, *Assembly of mutant-template telomerase RNA into catalytically active telomerase ribonucleoprotein that can act on telomeres is required for apoptosis and cell cycle arrest in human cancer cells.* Cancer Res, 2006. **66**(11): p. 5763-71.

237. Li, S., et al., *Rapid inhibition of cancer cell growth induced by lentiviral delivery and expression of mutant-template telomerase RNA and anti-telomerase short-interfering RNA.* Cancer Res, 2004. **64**(14): p. 4833-40.

238. Stohr, B.A. and E.H. Blackburn, *ATM mediates cytotoxicity of a mutant telomerase RNA in human cancer cells.* Cancer Res, 2008. **68**(13): p. 5309-17.

239. Guiducci, C., M.A. Cerone, and S. Bacchetti, *Expression of mutant telomerase in immortal telomerase-negative human cells results in cell cycle deregulation, nuclear and chromosomal abnormalities and rapid loss of viability.* Oncogene, 2001. **20**(6): p. 714-25.

240. Marusic, L., et al., *Reprogramming of telomerase by expression of mutant telomerase RNA template in human cells leads to altered telomeres that correlate with reduced cell viability.* Mol Cell Biol, 1997. **17**(11): p. 6394-401.

241. de Lange, T., *T-loops and the origin of telomeres.* Nat Rev Mol Cell Biol, 2004. **5**(4): p. 323-9.

242. Carneiro T, K.L., Reis C, Borges V, Moser B, Nakamura T, Ferreira M, *Telomeres avoid end detection by severing the checkpoint signal transduction pathway* Nature, 2010. **467**: p. 228-232.

243. Diolaiti, M.E., et al., *In situ visualization of telomere elongation patterns in human cells.* Nucleic Acids Res, 2013. **41**(18): p. e176.

244. Elena, S.F. and R.E. Lenski, *Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.* Nat Rev Genet, 2003. **4**(6): p. 457-69.

245. Darwin, C., *On the Origin of Species by Means of Natural Selection*. First Edition ed. 1859, London: John Murray.

246. Lenski, R.E., et al., *Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. .* American Naturalist, 1991. **138**: p. 1315-1341.

247. Puchta, O., et al., *Network of epistatic interactions within a yeast snoRNA.* Science, 2016. **352**(6287): p. 840-4.

248. Sprouffske, K., et al., *Cancer in light of experimental evolution.* Curr Biol, 2012. **22**(17): p. R762-71.